April 1993

LIDS-P-2175

Efficient Multiscale Regularization with Applications to the Computation of Optical Flow[1]

Mark R. Luettgen
W. Clem Karl
Alan S. Willsky

Department of EECS, MIT
Room 35-439
Cambridge, MA 02139
e-mail: luettgen@athena.mit.edu
tel: 617-253-6172
fax: 617-258-8553

Abstract

A new approach to regularization methods for image processing is introduced and developed using as a vehicle the problem of computing dense optical flow fields in an image sequence. Standard formulations of this problem require the computationally intensive solution of an elliptic partial differential equation which arises from the often used "smoothness constraint" type regularization. We utilize the interpretation of the smoothness constraint as a "fractal prior" to motivate regularization based on a recently introduced class of multiscale stochastic models. The solution of the new problem formulation is computed with an efficient multiscale algorithm. Experiments on several image sequences demonstrate the substantial computational savings that can be achieved due to the fact that the algorithm is non-iterative and in fact has a per pixel computational complexity which is independent of image size. The new approach also has a number of other important advantages. Specifically, multiresolution flow field estimates are available, allowing great flexibility in dealing with the tradeoff between resolution and accuracy. Multiscale error covariance information is also available, which is of considerable use in assessing the accuracy of the estimates. In particular, these error statistics can be used as the basis for a rational procedure for determining the spatially-varying optimal reconstruction resolution. Furthermore, if there are compelling reasons to insist upon a standard smoothness constraint, our algorithm provides an excellent initialization for the iterative algorithms associated with the smoothness constraint problem formulation. Finally, the usefulness of our approach should extend to a wide variety of ill-posed inverse problems in which variational techniques seeking a "smooth" solution are generally used.

EDICS category 1.11.

[1]This work was performed in part during a visit to the Institut de Recherche en Informatique et Systemes Aleatoires (IRISA), and was also supported by the Air Force Office of Scientific Research under Grant AFOSR-92-J-0002, by the Draper Laboratory IR&D Program under agreement DL-H-418524, by the Office of Naval Research under Grant N00014-91-J-1004 and by the Army Research Office under Grant DAAL03-92-G-0115.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
APR 1993		**00-04-1993 to 00-04-1993**

4. TITLE AND SUBTITLE
Efficient Multiscale Regularization with Applications to the Computation of Optical Flow

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Massachusetts Institute of Technology,Laboratory for Information and Decision Systems,77 Massachusetts Avenue,Cambridge,MA,02139

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)
11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**53**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

1 Introduction

In this paper we introduce and develop a new multiscale approach to regularization problems in image processing, using the computation of dense optical flow fields as the vehicle for our development. Regularization is, of course, a widely-known and used concept in image analysis. In some cases the introduction of a regularizing term is necessitated by ill-posedness (also referred to as the "aperture problem" in computer vision), i.e. by the insufficient information provided solely by the available data, or by a desire to reduce noise. In other problems the so-called regularizing term represents substantive prior information arising, for example, from physical constraints or laws or from information extracted from previous image frames. The family of optical flow reconstruction algorithms stemming from the work of Horn and Schunck [19], which forms the specific context for our development and which has found success in a number of applications such as [32], is one example of a formulation typically introduced to deal with ill-posedness. However, very similar formulations arise in other contexts ranging from the problem of the temporal tracking of optical flow [8] to large scale oceanographic data assimilation problems [36]. Thus, while we use the problem of estimating optical flow at a single point in time as the focus for our development, it is our strong belief that the ideas developed here have a far broader range of applicability.

Optical flow, the apparent velocity vector field corresponding to the observed motion of intensity patterns in successive image frames, is an important quantity in a variety of problems. For example, in MRI imaging of the heart [32, 30] this vector field provides diagnostic information concerning cardiac muscle motion and differential strain. In oceanographic data processing such information can be of use, for example, in tracking the meandering motion of the Gulf Stream [25]. Also, in computational vision, optical flow is an important input into higher level vision algorithms performing tasks such as segmentation, tracking, object detection, robot guidance and recovery of shape information [1, 27, 33, 37, 43]. In addition, methods for computing optical flow are an essential part of motion compensated coding schemes [2, 50].

As we have indicated, our approach to optical flow estimation is motivated by, and represents an alternative to, regularization methods such as that of Horn and Schunck [19] which employs the often used "smoothness constraint" regularization term. In particular, the starting point for this and many other approaches to optical flow estimation is the use of a brightness constraint, i.e. the assumption that changes in image brightness are due only to motion in the image frame. This leads to the so called *brightness constraint equation*[2] [19]:

$$0 = \frac{d}{dt}E(z_1, z_2, t) = \frac{\partial}{\partial t}E(z_1, z_2, t) + \nabla E(z_1, z_2, t) \cdot \boldsymbol{x}(z_1, z_2, t) \tag{1}$$

where $E(z_1, z_2, t)$ is the image intensity as a function of time t and space (z_1, z_2), $\boldsymbol{x}(z_1, z_2, t)$ is the optical flow vector field, and:

$$\boldsymbol{x} = \left[\frac{\partial z_1}{\partial t} \quad \frac{\partial z_2}{\partial t}\right]^T \tag{2}$$

$$\nabla E = \left[\frac{\partial E}{\partial z_1} \quad \frac{\partial E}{\partial z_2}\right] \tag{3}$$

[2] More generally, it is straightforward to adapt (1) to cases in which E has a known temporal variation. See [32] for an example in the context of MRI imaging.

The brightness constraint equation (1), however, does not completely specify the flow field $x(z_1, z_2, t)$ since it provides only one linear constraint for the two unknowns at each point. Thus, (1) by itself represents an under-determined or ill-posed set of constraints on optical flow. In addition, in practice, only noisy measurements of the temporal and spatial intensity derivatives will be available, meaning that we in fact have available only noisy constraints. For both of these reasons one must regularize the problem of reconstructing $x(z_1, z_2, t)$, and one commonly used way to do this is to assume some type of spatial coherence in the optical flow field, for instance by assuming that $x(z_1, z_2, t)$ is constant over spatial patches or by other methods for imposing coherence and achieving spatial noise averaging.

In particular, Horn and Schunck's approach [19], often referred to as imposing a *smoothness constraint*, consists of constructing the optical flow field estimate as the solution of the following optimization problem:

$$\hat{x}_{SC} = \arg\min_{x} \iint R^{-1} \left(\frac{d}{dt}E\right)^2 + \|\nabla x\|^2 \, dz_1 dz_2 \tag{4}$$

The smoothness constraint is captured by the second term which penalizes large gradients in the optical flow. The constant R allows one to tradeoff between the relative importance in the cost function of the brightness and smoothness constraint terms. For example, in some situations R^{-1} is taken to be quite large to force the solution to match the constraints (1), and in such a case the smoothness constraint serves merely to regularize the problem, i.e. to ensure that (4) has a unique solution. In other cases, however, one might use a more moderate value of R^{-1} either to account for the fact that the constraint (1) is noisy or to reflect the fact that the smoothness constraint penalty represents a useful source of information itself. For example, in [8] the smoothness constraint is replaced by an analogous term reflecting both smoothness and prior information gleaned from preceding image frames. We refer to the optical flow estimate obtained from (4) as the smoothness constraint (SC) solution to the problem of computing optical flow.

One of the major problems associated with the formulation in (4) and with analogous formulations for other regularized image processing problems is that they lead to computationally intensive algorithms. Specifically, one can show that the solution of (4) satisfies an elliptic partial differential equation (PDE) [19]. Discretization of this PDE leads to a sparse but extremely large set of linear equations which are typically solved using iterative approaches. One of the first iterative approaches used was the Gauss-Seidel relaxation algorithm [19, 40] which is extremely simple, but which converges very slowly. Terzopoulos [45] proposed the use of *multigrid* approaches and reported a factor of 7 reduction in computation over the Gauss-Seidel approach. Successive over-relaxation (SOR) algorithms [21] also provide significant computational improvement over GS approaches and have been successfully used in [32, 34, 35]. However, whatever numerical method is employed, computational complexity per pixel typically grows with image size, a fact that can make real-time or in some cases even off-line implementation prohibitively complex. For example, while computational complexity for such a problem may be severe for 512×512 images, especially if real-time processing of image sequences is required, the computational demands in other contexts, such as oceanographic data processing where one may consider problems as large as 100,000,000 voxels (3-D pixels), are more than a serious problem: they are, in fact, the *major* problem.

One of the principal motivations for the method in this paper is to introduce an alternative regularization formulation in order to address the computational challenge discussed above. To do this, we need to analyze the smoothness constraint in more detail. Note in particular that the penalty associated with the smoothness constraint term in (4) is equal to the integral of the squared norm of the field gradient over the image plane. In a one-dimensional context, such a constraint would penalize each of the (one-dimensional) fields in Figure 1 equally. Intuitively, the smoothness constraint has a fractal nature, and in fact is often referred to as a "fractal prior" [44].

Moreover, as discussed in [34, 35] and as described in more detail in the next section, the optical flow problem formulation in (4) has an equivalent formulation and precise interpretation in an estimation-theoretic context. Roughly speaking, the optimization problem (4) corresponds to a statistical model in which the noise or error in the brightness constraint is assumed to be spatially white and in which the two components of the optical flow are modeled as independent random fields, each of which has a zero mean, spatially white gradient. That is, as discussed in [8, 34, 35], the smoothness constraint essentially corresponds to modeling each component of optical flow as a spatial Brownian motion, i.e. as a statistically self-similar, fractal process with a $1/|\mathbf{f}|^2$ generalized spectrum [44].

Given that the smoothness constraint corresponds precisely to a prior model with fractal characteristics, a natural idea is that of using alternate prior statistical models — corresponding to alternate penalty terms to that in (4) — that possess the same type of fractal characteristics but that lead to computationally more attractive problem formulations. In this paper, we do just that as we introduce an approach based on substituting a fractal-like class of prior models recently introduced in [11, 9, 10, 13] for the smoothness constraint prior. The key idea behind this approach is that instead of the Brownian motion fractal prior that describes the optical flow field as one that has independent increments in space, we use a statistical model for optical flow that has independent increments in *scale*. That is, as described in the next section, we make use of a new class of statistical models for random fields that describe these fields in a scale-recursive manner, with detail added as we move from coarse-to-fine scales. The model can be interpreted as a smoothness constraint that provides individual penalties on each scale of detail or as providing a multiscale probabilistic model in which the variances of the detail components vary from scale to scale in a fractal, self-similar fashion. For this reason, we say that our formulation corresponds to a multiscale regularization (MR) of the optical flow problem, and we refer below to the MR algorithm and solution.

One of the most important consequences of this alternate smoothness constraint is that it allows us to make use of the extremely efficient scale-recursive optimal estimation algorithm that this statistical model admits [11, 9, 10]. In particular, the resulting algorithm is not iterative and in fact requires a fixed number of floating point operations per pixel *independent of image size*. Thus, since methods for solving the smoothness constraint problem formulation have per pixel computational complexities that typically grow with image size, the computational savings associated with the new approach increases as the image size grows and, as we will see, can be considerable even for modest-sized problems.

Moreover, while computational efficiency did serve as the original motivation for this new formulation and in many problems may be its most important asset, there are several other potential advantages that the new approach has. First, the scale-recursive nature

of our algorithm directly yields estimates of the optical flow field at multiple resolutions, providing us with considerable flexibility in dealing with the tradeoff between accuracy and resolution. Specifically, one can expect to obtain higher accuracy at coarser resolutions, and thus one can imagine trading off resolution versus accuracy in a data-adaptive way. For example, in regions with substantial local intensity variations one would expect to be able to estimate optical flow at a finer spatial resolution than in regions in which intensity varies more smoothly and contrast is low. The question, of course, is how such an intuitive concept can be realized in an algorithm. As we will demonstrate, our multiscale algorithm provides us with all of the information required to do this with essentially no additional computation, leading to the designation of the preferred resolution for estimating optical flow at every point in the image frame.

Secondly, an important consequence of employing an estimation-theoretic interpretation is that it offers the possibility of evaluating a quantitative measure of the quality of our optical flow estimate, namely the estimation error covariance. This idea, of course, also applies to the original smoothness constraint formulation (4). However, in that case, the computation of the error covariance must be done in addition to solving the partial differential equations for the optimal flow estimates, and in fact, the computation of these error statistics has complexity at least as great as that for calculating the estimates. In contrast, for our formulation, error covariances can be calculated with essentially no increase in computational complexity. Furthermore, our algorithm provides error covariance statistics at multiple resolutions, providing information that is essential to addressing the tradeoff between resolution and accuracy as discussed in the previous paragraph, and that may also be useful to higher level vision algorithms which need to combine information in a rational way from a variety of sources [38].

As we have indicated, the new algorithm we develop is based on a formulation that is similar but not identical to that given by (4), and there are several implications of this fact. The first is that while the estimates produced by our algorithms are not identical to those based on (4), they are similar and have comparable root-mean-square (rms) error characteristics, as the experimental evidence in Section 3 illustrates. Moreover, these results also show that the difference between the SC and MR flow estimates consists of mostly high spatial frequency components, which are precisely the components which can be quickly removed by the iterative algorithms computing a smoothness constraint solution. Thus, even in situations in which a solution to the original smoothness constraint formulation is required (for instance, if the smoothness constraint corresponds to physically-based prior information) there may be considerable computational advantages in using the MR solution as an initial estimate of the optical flow, i.e. as an initial estimate for an iterative algorithm which computes the solution of the partial differential equation characterizing (4). Indeed, given the promise suggested by results presented here, we conjecture that another potential application of the approach we introduce is in providing easily computed, accurate initial conditions for the solution of partial differential equations arising in contexts other than image processing.

There is another implication of the relationship of our approach to the formulation in (4). Specifically, there are of course, problems of practical importance in which the basic assumptions underlying the Horn and Schunck formalism are violated, for instance if there is substantial temporal aliasing (so that the data implied by (1) are not available), if there

are discontinuities in the motion field due to object boundaries and occlusion or if there are multiple motions. In such cases, the Horn and Schunck formulation may fail to give adequate results, and, due to the similarity of the approaches, our method would likely fail as well. In such contexts algorithms developed to deal explicitly with such issues, such as those in [15, 18], may be more appropriate. On the other hand, for the not insignificant class of problems for which the Horn and Schunck formulation is well-suited, such as [32] and the many ill-posed and variational problems arising in fields ranging from image processing and tomography to meteorology, seismology and oceanography [5, 31, 47, 22, 46, 26], our method will also work well and also provides the advantages described previously: computational efficiency, multiresolution estimates and multiscale error covariance information. Moreover, even in cases in which Horn and Schunck-type global smoothness constraints are inappropriate, there are reasons to believe that algorithms based on our formulation may provide the basis for promising new solutions. While it is beyond the scope of this paper to develop such methods in detail, we provide an example suggesting this promise and also indicate how the statistical interpretation and flexible structure of our formalism might be used to advantage.

This paper is organized as follows. In Section 2 we discuss in more detail an estimation-theoretic interpretation of the optical flow formulation in (4) and develop our new approach to the computation of optical flow. Section 3 presents experimental results on several real and synthetic image sequences. Section 4 provides further discussion and conclusions.

2 Multiscale Regularization

In the first part of this section we develop a discrete formulation of the optical flow problem, and discuss in more detail the estimation-theoretic interpretation of it. We then illustrate precisely how the smoothness constraint can be interpreted as a prior model for the flow field, and how it can be replaced by another, similar prior model which leads to a more computationally attractive problem formulation. The general class of prior models we use is then introduced along with an algorithm for finding the solution of the new optical flow problem formulation.

2.1 An Estimation-Theoretic Interpretation of the Optical Flow Problem

Estimation-theoretic formulations and interpretations of optical flow problems have been introduced and studied by a number of authors. For instance, in [20, 49] Markov random field (MRF) models are proposed along with a *maximum a-posteriori* criterion for estimating optical flow. MRF models are also used in [18] to address problems of occlusion and flow field discontinuity. Kalman filtering approaches which allow for temporal as well as spatial smoothness constraints have been discussed in [8, 39, 17, 42]. In addition, in [38] a Bayesian formulation which provides optical flow estimates and confidence measures based on a local window of data is proposed. In addition there is the interpretation by Rougée et al. [34, 35] of the Horn and Schunck smoothness constraint formulation (4) as an equivalent estimation problem with a Brownian motion, fractal prior for the flow field. The distinguishing feature of the Brownian motion model implied by (4), the Markov random field models, and the spatio-temporal models used in the Kalman filtering approaches, is that

they all provide models in terms of local relationships (typically nearest neighbor) of the flow field components at a single, finest level of resolution. This leads naturally to spatially local, iterative algorithms for computing the optimal optical flow estimates (such as those needed to solve the partial differential equation resulting from (4) or simulated annealing algorithms for MRF models). In contrast, the probabilistic model for optical flow proposed in this paper describes the flow field in terms of probabilistic variations from scale to scale and leads naturally to the efficient scale recursive algorithms described in [11, 9, 10].

As we have indicated, our approach is motivated by the probabilistic interpretations of Horn and Schunck's formulation, which we now discuss briefly. The reader referred to [7, 8, 34, 35] for a more extensive discussion of this and related probabilistic models. We start by introducing the following notation. Define:

$$y(z_1, z_2) \equiv -\frac{\partial}{\partial t} E(z_1, z_2, t) \tag{5}$$

$$C(z_1, z_2) \equiv \nabla E(z_1, z_2, t) \tag{6}$$

The brightness constraint equation (1) can then be written:

$$y(z_1, z_2) = C(z_1, z_2) \cdot \boldsymbol{x}(z_1, z_2) \tag{7}$$

where the time dependence of the equations has been suppressed.

In practice, brightness measurements are only available over a discrete set of points in space and time. Thus, the temporal and spatial derivative terms in the brightness constraint equation (7) must be approximated by a finite difference scheme in time and space, and the optical flow is only estimated on a discrete space-time grid. There are a number of important issues which arise due to the discretization, such as the use of spatial and/or temporal smoothing prior to discretization, the use of more than two image frames in the computation of temporal derivatives, etc., and we refer the reader to [7, 3, 15] for further discussion. We assume here that the optical flow is to be estimated on the set $\{(z_1, z_2) | z_1 = ih, z_2 = jh; i, j \in \{1, \cdots, 2^M\}\}$ where h is the grid spacing and M is an integer. The assumption that the lattice is square and that the number of rows is equal to a power of two simplifies the notation in the subsequent development, but is not essential as we discuss in Appendix A. In order to simplify the notation further, we let $y(i,j)$, $\boldsymbol{x}(i,j)$, and $C(i,j)$ denote the measured temporal brightness derivative, the optical flow, and the spatial gradient of the image brightness, respectively, at grid point (ih, jh). The brightness constraints at all grid points can then be grouped into one large set of linear equations to capture the optical flow information contained in the image sequence. Defining $\mathbf{x}$ as the vector of optical flow vectors $\boldsymbol{x}(i,j)$ at all grid points (using, say, a lexicographic ordering), $\mathbf{C}$ as the matrix containing the corresponding spatial gradient terms $C(i,j)$, and $\mathbf{y}$ as the vector of temporal gradients $y(i,j)$, we can write:

$$\mathbf{y} = \mathbf{C}\mathbf{x} \tag{8}$$

Then, the discrete counterpart of (4) is:

$$\begin{aligned} \hat{\mathbf{x}}_{SC} &\equiv \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{C}\mathbf{x}\|^2_{\mathbf{R}^{-1}} + \|\mathbf{L}\mathbf{x}\|^2_{\mathbf{I}} \\ &= \arg\min_{\mathbf{x}} (\mathbf{y} - \mathbf{C}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{C}\mathbf{x}) + \mathbf{x}^T \mathbf{L}^T \mathbf{L}\mathbf{x} \end{aligned} \tag{9}$$

where the matrix $\mathbf{L}$ is a discrete approximation of the gradient operator in (4) and $\mathbf{R} = R\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. The regularization term $\mathbf{x}^T\mathbf{L}^T\mathbf{L}\mathbf{x}$ makes the optimization problem (9) well-posed. In particular, the solution of (13) satisfies the so-called *normal equations* [41]:

$$(\mathbf{C}^T\mathbf{R}^{-1}\mathbf{C} + \mathbf{L}^T\mathbf{L})\hat{\mathbf{x}}_{SC} = \mathbf{C}^T\mathbf{R}^{-1}\mathbf{y} \tag{10}$$

and the invertibility of $(\mathbf{C}^T\mathbf{R}^{-1}\mathbf{C} + \mathbf{L}^T\mathbf{L})$ guarantees that $\hat{\mathbf{x}}_{SC}$ is unique. The normal equations (10) are the discrete counterpart of the partial differential equation that arises from (4).

An estimation-theoretic formulation of the optimization problem in (9) can now be developed. Specifically, suppose that we wish to estimate $\mathbf{x}$ based on the measurements

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{v} \tag{11}$$

$$\mathbf{0} = \mathbf{L}\mathbf{x} + \mathbf{w} \tag{12}$$

where $\mathbf{v}$ and $\mathbf{w}$ are uncorrelated random vectors with[3] $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Then the measurement vector $\bar{\mathbf{y}} \equiv [\mathbf{y}^T|\mathbf{0}]^T$ is conditionally Gaussian, and the maximum likelihood estimate [48] of $\mathbf{x}$ is:

$$\begin{aligned}
\hat{\mathbf{x}}_{ML} &\equiv \arg\max_{\mathbf{x}} p(\bar{\mathbf{y}}|\mathbf{x}) \\
&= \arg\min_{\mathbf{x}} -\log p(\bar{\mathbf{y}}|\mathbf{x}) \\
&= \arg\min_{\mathbf{x}} (\mathbf{y} - \mathbf{C}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{C}\mathbf{x}) + \mathbf{x}^T\mathbf{L}^T\mathbf{L}\mathbf{x} \\
&= \hat{\mathbf{x}}_{SC}
\end{aligned} \tag{13}$$

Thus, the maximum likelihood problem formulation results in the same solution as the smoothness constraint formulation when $\mathbf{L}$ is used to define an additional set of noisy measurements.

The main point here is that by formulating the problem in this estimation-theoretic framework, we can use (12) to interpret the smoothness constraint as a prior probabilistic model for the flow field. Specifically, we can rewrite (12) as:

$$\mathbf{L}\mathbf{x} = -\mathbf{w} \tag{14}$$

Recalling that $\mathbf{L}$ is an approximation to the gradient operator, we see that (14) is nothing more than a spatial difference equation model for $\mathbf{x}$ driven by the spatial white noise field $\mathbf{w}$.

To some extent the precise form of this prior model is arbitrary, and thus we are led to the idea of introducing a new prior model which is similar in nature, but which leads to a computationally more attractive problem formulation. That is, we want to change the smoothness constraint term $\mathbf{x}^T\mathbf{L}^T\mathbf{L}\mathbf{x}$ in (13) to something similar, say, $\mathbf{x}^T\mathbf{S}\mathbf{x} \approx \mathbf{x}^T\mathbf{L}^T\mathbf{L}\mathbf{x}$ (where $\mathbf{S}$ is a symmetric positive semi-definite matrix) such that the resulting optimization problem is easy to solve. If we factor $\mathbf{S}$ as $\mathbf{S} = \bar{\mathbf{L}}^T\bar{\mathbf{L}}$ then we can interpret the new constraint as a prior probabilistic model just as we did with the smoothness constraint. In addition,

[3]The notation $\mathbf{z} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Lambda})$ means that $\mathbf{z}$ has a Gaussian distribution, with mean $\mathbf{m}$ and variance $\mathbf{\Lambda}$.

there is a precise interpretation of what we have done as a Bayesian estimation problem. Specifically, if $\mathbf{S}$ is invertible, then the use of this new constraint in place of the smoothness constraint is equivalent to modeling the flow field probabilistically as $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{S}^{-1})$, since in this case the Bayes' least squares estimate of the flow field $\mathbf{x}$, given this prior model and the measurements in (11) is given by:

$$\hat{\mathbf{x}}_{BLSE} = \arg\min_{\mathbf{x}} (\mathbf{y} - \mathbf{C}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{C}\mathbf{x}) + \mathbf{x}^T \mathbf{S} \mathbf{x} \tag{15}$$

which corresponds to (13) with a different prior model term. The normal equations corresponding to (15) are given by:

$$(\mathbf{C}^T \mathbf{R}^{-1} \mathbf{C} + \mathbf{S}) \hat{\mathbf{x}}_{BLSE} = \mathbf{C}^T \mathbf{R}^{-1} \mathbf{y} \tag{16}$$

Comparison of the problem formulations (9) and (15), or of the normal equations (10) and (16), makes it apparent how the two problem formulations are related. Note that an analogous Bayesian interpretation can apparently be given to the smoothness constraint formulation (9), (10), with the corresponding prior model for optical flow given by $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, (\mathbf{L}^T \mathbf{L})^{-1})$. Recall, however, that $\mathbf{L}$ is an approximation to the spatial gradient operator and thus is *not* invertible since operating on constants with this operator yields zero. The probabilistic interpretation of this is that the model (14) places probabilistic constraints on the spatial *differences* of the optical flow, but not on its DC value. Indeed, it is not difficult to check that if we model optical flow instead as $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, (\mathbf{L}^T \mathbf{L} + \epsilon I)^{-1})$, where ϵ is any arbitrarily small positive number, then $\mathbf{L}^T \mathbf{L} + \epsilon I$ is indeed invertible and the DC value of $\mathbf{x}$ has a prior covariance P_0 on the order of $1/\epsilon$, so that $P_0 \to \infty$ as $\epsilon \to 0$. Thus, the original smoothness constraint formulation in essence assumes an infinite prior covariance on the DC value of optical flow. The alternate model developed in the next section has a similar parameter, P_0, representing the DC variance, which can similarly be set to ∞.

Finally, it is important to emphasize that what we have done here is to *interpret* the smoothness constraint formulation and its extension (15) as optimal estimation problems. The point is that we are not assuming statistics for $\mathbf{x}$ and $\mathbf{v}$ but rather are identifying the assumptions that are intrinsic to the smoothness constraint formulation. That is, $\hat{\mathbf{x}}_{BLSE}$ in (15) *is* the Bayes' least squares estimate if $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{S}^{-1})$ and $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. More generally, if $\mathbf{x}$ and $\mathbf{v}$ are simply modeled as zero-mean uncorrelated random vectors with covariances $\mathbf{S}^{-1}$ and $\mathbf{R}$, respectively, and with no further specification of their distributions, then (15) is the linear least squares estimate, i.e. the best linear estimate of $\mathbf{x}$.

The choice of the new prior model is now clearly at the heart of the problem. Recalling that the smoothness constraint has the interpretation as a "fractal prior", we choose a prior model which also has fractal-like characteristics. A natural way to specify such models is to explicitly represent the optical flow field at multiple scales so that the self-similar fractal characteristics of the field can be introduced explicitly. A stochastic modeling framework which allows us to do this, and which also leads to efficient algorithms for solving (15), (16), is described in the next section.

2.2 A Class of Multiscale Models

The models we utilize to replace the smoothness constraint prior model were recently introduced in [11, 9, 10, 13]. The models represent the flow field at multiple scales, i.e. for

a set of scales $m = 0, \ldots, M$, with $m = 0$ being the coarsest scale and $m = M$ the finest scale, we define a set of optical flow fields indexed by scale and space, namely $x_m(i,j)$. At the m^{th} scale, the field consists of 4^m flow vectors, as illustrated in Figure 2, capturing features of the optical flow field discernible at that scale (i.e. finer resolution features of the field appear only in finer scale representations). Thus, the coarsest version of the flow field consists of just a single vector corresponding to the coarse, aggregate value of the optical flow over the entire spatial domain of interest, and successively finer versions consist of a geometrically increasing number of vectors. At the finest level, the flow field is represented on a grid with the same resolution as the image brightness data. In particular, $x_M(i,j)$ corresponds to the optical flow vector $x(i,j)$.

The multiscale optical flow field is defined on the *quadtree structure* illustrated in Figure 3. Pyramidal data structures such as the quadtree naturally arise in image processing algorithms which have a multiresolution component. For instance, successive filtering and decimation operations lead to images defined on such a hierarchy of grids in the Laplacian pyramid coding algorithm of Burt and Adelson [6] and in the closely related wavelet transform decomposition of images [23]. Also, the multigrid approaches to low level vision problems discussed by Terzopoulos [45] involve relaxation on a similar sequence of grids. It is important to emphasize here, however that in contrast to approaches such as these, in our case we are using the quadtree structure to *model* a spatially-distributed random field rather than to analyze or decompose a given field. As we will see, this model does, in fact, lead to processing algorithms operating on the quadtree, but these algorithms are optimal estimation procedures and thus are completely different in form, nature, and intent from standard pyramidal decomposition procedures.

Our quadtree model for the optical flow field $x(i,j) = x_M(i,j)$ is constructed by adding detail from one scale to the next (i.e. from coarse to fine). Just as the smoothness constraint prior model (14) describes probabilistic constraints among values of the optical flow at different spatial locations, our multiscale model describes such constraints among values at different *scales*. For notational convenience in describing such models, we denote nodes on the quadtree with a single abstract index s which is associated with the 3-tuple (m,i,j) where, again, m is the scale and (i,j) is a spatial location in the grid at the m^{th} scale (see Figure 2). It is also useful to define an *upward shift operator* $\bar{\gamma}$. In particular, the *parent* of node s is denoted $s\bar{\gamma}$ (see Figure 3). For instance, if s corresponds to any of the nodes in the upper left quadrant of the second level grid (see Figure 2), i.e. nodes $(2,1,1),(2,2,1),(2,1,2)$ or $(2,2,2)$, then $s\bar{\gamma}$ corresponds to their parent on the first level, namely node $(1,1,1)$. With this notation, our scale-recursive model takes the form:

$$x(s) = A(s)x(s\bar{\gamma}) + B(s)w(s) \tag{17}$$

under the following assumptions:

$$x_0 \sim \mathcal{N}(0, P_0) \tag{18}$$

$$w(s) \sim \mathcal{N}(0, I) \tag{19}$$

The vectors x and w are referred to as the state and driving noise terms. The state variable x_0 at the root node of the tree provides an initial condition for the recursion. The driving noise is white in both space and scale, and is uncorrelated with the initial condition. Interpreting each level as a representation of a two-dimensional field, we see that

(17) describes the evolution of the process from coarse to fine scales. The term $A(s)x(s\bar{\gamma})$ represents interpolation down to the next level, and $B(s)w(s)$ represents higher resolution detail added as the process evolves from one scale to the next. In the application of interest here, $x(s) = x_m(i,j)$, where $s = (m,i,j)$, and thus $A, B \in \Re^{2\times 2}$. Such a model corresponds in essence to a first-order recursion in scale for optical flow.[4]

Measurements of the finest level optical flow field are available from the brightness constraint. In particular, at a particular point (i,j) at the finest level M, we have the measurement equation:

$$y(i,j) = C(i,j)x_M(i,j) + v(i,j) \tag{20}$$
$$v(i,j) \sim \mathcal{N}(0,R) \tag{21}$$

where $C(i,j) \in \Re^{1\times 2}$ and the white Gaussian observation noise is assumed to be independent of the initial condition x_0 and the driving noise w in (17) – (19). Of course, we can group the state variables $x(s)$ at the finest level into a vector $\mathbf{x}_M$ as well as the corresponding measurements $y(s)$ and spatial gradient terms $C(s)$ in the same way as we did to get (8):

$$\mathbf{y} = \mathbf{C}\mathbf{x}_M + \mathbf{v} \tag{22}$$
$$\mathbf{v} \sim \mathcal{N}(\mathbf{0},\mathbf{R}) \tag{23}$$

We now have exactly the framework which led to the statement of (15) as a generalization of the smoothness constraint formulation (13). In particular, the modeling equations (17) – (19) indicate that at the finest level of the quadtree, the flow field vectors will be a set of jointly Gaussian random variables $\mathbf{x}_M \sim \mathcal{N}(\mathbf{0},\mathbf{\Lambda})$, where $\mathbf{\Lambda}$ is implicitly given by the parameters in (17) – (19), and a set of noisy measurements given by (22). The Bayes' least squares estimate of $\mathbf{x}_M$ given the measurements in (22) and the prior model (17) – (19) is:

$$\hat{\mathbf{x}}_M = \arg\min_{\mathbf{x}_M} (\mathbf{y} - \mathbf{C}\mathbf{x}_M)^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{C}\mathbf{x}_M) + \mathbf{x}_M^T \mathbf{\Lambda}^{-1} \mathbf{x}_M \tag{24}$$

The multiscale modeling framework thus provides an alternative to the smoothness constraint formulation of (9) or (13). Furthermore, if we drop the assumption of Gaussianity for $x_0, w(s)$, and $v(i,j)$, the optimal estimate $\hat{\mathbf{x}}_M$ has the interpretation as the linear least squares estimate of $\mathbf{x}$.

What remains to be done are (1) to specify a model within this class that has characteristics similar to those of the smoothness constraint prior model, and (2) to demonstrate why the use of this alternate multiresolution formulation is of any interest. We defer the latter of these to the next section and focus here on the former. In particular, for our multiscale model based on (17) – (19) to approximate the smoothness constraint prior we would like to choose our model parameters so that we have $\mathbf{\Lambda}^{-1} \approx \mathbf{L}^T\mathbf{L}$. The observation that the prior model implied by the operator $\mathbf{L}$ in (13) corresponds to a Brownian motion "fractal prior" suggests one approach to choosing the model parameters. In particular,

[4] More generally, higher-order recursions in scale can be captured, just as in standard state space models, by increasing the order of the model, i.e. the dimension of $x(s)$. In this case the actual optical flow at node s would correspond to a subset of the components of $x(s)$, with the remainder of $x(s)$ devoted to capturing the memory in the multiscale recursion. In this paper, however, we restrict ourselves to the simple first order recursion.

the one-dimensional Brownian motion has a $1/f^2$ generalized spectrum [24]. It has been demonstrated that such processes are well approximated by multiscale models such as ours in one dimension if geometrically decreasing powers of noise are added at each level m of the process [10, 52]. This motivates the choice of $B(s) = b4^{-\frac{\mu m(s)}{2}} I$ in (17), where I is the 2×2 identity matrix, and where b and μ are scalar constants. The constant b directly controls the overall noise power in the process. Also, as discussed in [52], the choice of μ controls the power law dependence of the generalized spectrum of the process at the finest resolution as well as the fractal dimension of its sample paths. Specifically, this spectrum has a $1/f^{\mu}$ dependence and the choice of $\mu = 2$ would correspond to a Brownian-like fractal process. Thus, our model for the optical flow field can be interpreted as providing individual penalties on each scale of detail, with penalty weights that vary from scale-to-scale in essentially the same way as the smoothness constraint's.

To achieve greater flexibility in both the modeling and estimation, we allow μ to be a parameter that can be varied. In addition, recall that in the smoothness constraint formulation, $\mathbf{L}^T\mathbf{L}$ was not invertible because of the implicit assumption of infinite prior variance on the DC value of the optical flow field. In our multiscale regularization context, this would correspond to setting P_0 equal to infinity in (18). This can be done without difficulty in the estimation algorithms described next, but we have found that it is generally sufficient simply to choose P_0 to be a large multiple of the identity.

2.3 The Multiscale Regularization Algorithm

We have now specified a class of models which will allow us to approximate the smoothness constraint prior model. The simple multiscale structure of these models leads to very efficient algorithms for computing the optimal estimate of the state given a set of measurements. One of these algorithms, which we refer to as the Multiscale Regularization (MR) algorithm, was developed in [11, 9, 10, 12] for one-dimensional signals, and its extension to images is described here.

The MR algorithm computes the Bayes' least squares estimate of the state vectors (17) given the measurements (20) in two steps. The first step is an *upward* or *fine-to-coarse* sweep on the quadtree, which propagates the measurement information in parallel, level by level, from the fine scale nodes up to the root node. The second step is a *downward* or *coarse-to-fine* sweep which propagates the measurement information back down, and throughout the tree. The result is the least squares estimate of the state $x(s)$ at each node based on all of the data. The details of the upward and downward sweeps are given below and are discussed in much greater detail in [10, 12].

To begin, note first that the measurement model (20) can be written in the following form, allowing for the possibility of spatially varying noise intensity:

$$y(s) = C(s)x(s) + v(s) \tag{25}$$

$$v(s) \sim \mathcal{N}(0, R(s)) \tag{26}$$

In the context of the optical flow estimation problem, measurements are taken only on the finest level, corresponding to $C(s) = 0$ unless s is a node at the finest level. However, in the more general modeling framework discussed in [10, 12], the measurements may be available

at any node, and the noise variance may vary with node as in (26). We present here this more general algorithm in which, in addition, x, y and w may be of arbitrary dimension.

The model given by (17) - (19), (25) - (26) is a downward model in the sense that the recursion starts from the root node and propagates down the quadtree from coarse-to-fine scales. In order to describe the upward sweep of the MR algorithm, we need a corresponding *upward model.* This upward model is equivalent to the downward model in the sense that the joint second order statistics of the states $x(s)$ and measurements $y(s)$ are the same. The upward model is given by[5] [9, 10]:

$$x(s\bar{\gamma}) = F(s)x(s) - A^{-1}(s)B(s)\tilde{w}(s) \quad (27)$$

$$y(s) = C(s)x(s) + v(s) \quad (28)$$

where:

$$F(s) = P_{s\bar{\gamma}}A^T(s)P_s^{-1} \quad (29)$$

$$\tilde{w}(s) = w(s) - \mathbf{E}[w(s)|x(s)] \quad (30)$$

$$\mathbf{E}[\tilde{w}(s)\tilde{w}^T(s)] = I - B(s)^T P_s^{-1} B(s) \quad (31)$$

$$\equiv \tilde{Q}(s) \quad (32)$$

and where $P_s = \mathbf{E}[x(s)x^T(s)]$ is the variance of the state at node s and evolves according to the Lyapanov equation:

$$P_s = A(s)P_{s\bar{\gamma}}A^T(s) + B(s)B^T(s) \quad (33)$$

To proceed further we need to define some new notation.

$$Y_s = \{y(s')|s' = s \text{ or } s' \text{ is a descendant of } s\} \quad (34)$$

$$Y_s^+ = Y_s \setminus \{s\} \quad (35)$$

$$\widehat{x}(s'|s) = \mathbf{E}[x(s')|Y_s] \quad (36)$$

$$\widehat{x}(s'|s+) = \mathbf{E}[x(s')|Y_s^+] \quad (37)$$

$$\widehat{P}(s'|s) = \mathbf{E}[(x(s') - \widehat{x}(s'|s))(x(s') - \widehat{x}(s'|s))^T] \quad (38)$$

$$\widehat{P}(s'|s+) = \mathbf{E}[(x(s') - \widehat{x}(s'|s+))(x(s') - \widehat{x}(s'|s+))^T] \quad (39)$$

where the notation $Y_s \setminus \{s\}$ means the node s is not included in the set Y_s^+. The upward sweep of the MR algorithm begins with the initialization of $\widehat{x}(s|s+)$ and the corresponding error covariance $P(s|s+)$ at the finest level, i.e. for s of the form (M, i, j) where M is the finest scale. The initial conditions at this scale reflect the prior statistics of $x(s)$ at scale M, as we have not yet incorporated data. Thus, for every s at this finest scale we set $\widehat{x}(s|s+)$ to zero (which is the prior mean of $x(s)$) and similarly set $P(s|s+)$ to the corresponding covariance, namely the solution of the Lyapanov equation (33) at the finest level. The upward sweep of the MR algorithm then proceeds recursively. Specifically, suppose that we

[5] We use $\mathbf{E}[x]$ to denote the expected value of the random variable x and $\mathbf{E}[x|y]$ to denote the expected value of x given y.

have $\hat{x}(s|s+)$ and $P(s|s+)$ at a given node s. Then this estimate is *updated* to incorporate the measurement $y(s)$ (if there is a measurement at this node) according to the following:

$$\begin{aligned}
\hat{x}(s|s) &= \hat{x}(s|s+) + K(s)[y(s) - C(s)\hat{x}(s|s+)] && (40)\\
P(s|s) &= [I - K(s)C(s)]P(s|s+) && (41)\\
K(s) &= P(s|s+)C^T(s)V^{-1}(s) && (42)\\
V(s) &= C(s)P(s|s+)C^T(s) + R(s) && (43)
\end{aligned}$$

Denote the offspring of $x(s)$ as $x(s\alpha_i), i = 1, \cdots, q$. For the quadtree model, of course, $q = 4$, but there is no increase in complexity here if we allow the possibility that there are more or fewer offspring for each node. The updated estimates at the offspring nodes are then *predicted* back up to the next level:

$$\begin{aligned}
\hat{x}(s|s\alpha_i) &= F(s\alpha_i)\hat{x}(s\alpha_i|s\alpha_i) && (44)\\
P(s|s\alpha_i) &= F(s\alpha_i)P(s\alpha_i|s\alpha_i)F^T(s\alpha_i) + Q(s\alpha_i) && (45)\\
Q(s\alpha_i) &= A^{-1}(s\alpha_i)B(s\alpha_i)\tilde{Q}(s\alpha_i)B^T(s\alpha_i)(A^{-1}(s\alpha_i))^T && (46)
\end{aligned}$$

The predicted estimates from the q offspring are then *merged*:

$$\begin{aligned}
\hat{x}(s|s+) &= P(s|s+)\sum_{i=1}^{q} P^{-1}(s|s\alpha_i)\hat{x}(s|s\alpha_i) && (47)\\
P(s|s+) &= [(1-q)P_s^{-1} + \sum_{i=1}^{q} P^{-1}(s|s\alpha_i)]^{-1} && (48)
\end{aligned}$$

The upward sweep given by the update, predict and merge equations proceeds recursively up the quadtree. At the top of the tree (corresponding to the root node $s = 0$), one obtains the *smoothed* estimate of the root node, that is, an estimate based on all of the data. The estimate and its error covariance are given by:

$$\begin{aligned}
\hat{x}^s(0) &= \hat{x}(0|0) && (49)\\
P^s(0) &= P(0|0) && (50)
\end{aligned}$$

where the superscript s denotes the fact that these are *smoothed* estimates. The smoothed estimate and associated error covariance at the root node provide initialization for the *downward sweep*, which is given by the following coarse-to-fine recursion:

$$\begin{aligned}
\hat{x}^s(s) &= \hat{x}(s|s) + J(s)[\hat{x}^s(s\bar{\gamma}) - \hat{x}(s\bar{\gamma}|s)] && (51)\\
P^s(s) &= P(s|s) + J(s)[P^s(s\bar{\gamma}) - P(s\bar{\gamma}|s)]J^T(s) && (52)\\
J(s) &= P(s|s)F^T(s)P^{-1}(s\bar{\gamma}|s) && (53)
\end{aligned}$$

The estimates $\hat{x}^s(s)$ at the finest level of the quadtree provide the solution to (24). The form of the algorithm we have specified here, which generalizes standard Kalman filtering and smoothing algorithms to the multiscale context, obviously assumes that the state covariance P_s is well defined and finite, and it is not difficult to see from (33) that this will be the case if P_0 is finite. There is, however, an alternate form of this algorithm presented in [10, 12]

which generalizes so-called *information form* algorithms for standard state space models and which propagates *inverses* of covariances. In this alternate form it is straightforward to accommodate the setting of P_0 to infinity (which corresponds to $P_0^{-1} = 0$), and we refer the reader to [10, 12] for details. As mentioned previously, we have found that setting P_0 to a large but finite multiple of the identity, and then using (40) – (48), (51) – (53), yields excellent results.

3 Experimental Results

3.1 Specification of the Multiscale Model

To specify the MR algorithm completely we use the following parameterization of the model (17) – (19), (25) – (26):

$$\begin{aligned} x(s) &= x(s\bar{\gamma}) + (b4^{\frac{-\mu m(s)}{2}})w(s) && (54)\\ y(s) &= C(s)x(s) + v(s) && (55)\\ w(s) &\sim \mathcal{N}(0, I) && (56)\\ v(s) &\sim \mathcal{N}(0, R(s)) && (57)\\ x_0 &\sim \mathcal{N}(0, pI) && (58) \end{aligned}$$

where I is a 2×2 identity matrix. From (54) and (56) we see that the two components of the optical flow field are modeled as independent sets of random variables, and that each has a fractal-like characteristic due to the form of the driving noise gain $B(s)$. The independence of the flow field components is motivated by the fact that the smoothness constraint formulation implicitly makes this assumption as well [34, 35]. We view μ and b as free model parameters which can be varied to control the degree and type of regularization in much the same way that the parameter R in the smoothness constraint formulation (4) is used to tradeoff between the data dependent and regularization terms in the optimization functional. However, we have found in our experiments that the choice $b = \mu = 1$ typically works well, and we have used these values in all of the experiments below.

As discussed previously, the measurements $y(s)$ and measurement matrix $C(s)$ come directly from the image temporal and spatial gradients, which are available at the finest level of the quadtree. In the experiments described below, we smoothed the images with the 7×7 filter given by:

$$\begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix} * \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix} * \cdots * \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix} \qquad (59)$$

(where $*$ denotes the 2-D convolution) and then calculated spatial gradients with a central difference approximation. The temporal gradient was computed as the difference of two smoothed images. Temporal smoothing (in addition to the spatial smoothing) has been shown to reduce estimation errors in several methods, including the smoothness constraint approach [3] and thus would be of value for the multiscale regularization method as well. For our purposes here, however, namely to demonstrate comparative computational efficiency relative to the smoothness constraint formulation and to illustrate the use and value of

both multiresolution estimates and covariance information, the simple two frame difference is sufficient.

The additive noise variance is given by $R(s)$. We have found empirically that the choice $R(s) = max(\|C(s)\|^2, 10)$ worked well in all cases. This choice effectively penalizes large spatial gradients, which are points at which the brightness constraint equation is likely to have large errors [38] (due, for example, to noise, aliasing or occlusion). The parameter p in the prior covariance (58) of the MR model root node was set to $p = 100$.

We compare our approach computationally and visually to the Gauss-Seidel (GS) and successive over-relaxation (SOR) algorithms, which can be used to compute the solution of the smoothness constraint formulation given by (9) or (13)(see, for example, [19, 21, 32, 34, 35, 40]. In our experiments, we have found that SOR typically provides a factor of 10 to 100 performance improvement of Gauss-Seidel, and hence is computationally equal to or better than multigrid approaches [45, 14]. The parameter R in the Horn and Schunck formulation (4) was chosen in to yield good visual and quantitative results. In particular, R was set to 100 in the first example below, and 2500 in the subsequent examples. Several possibilities for choosing this parameter based on the image data have been proposed in the literature [5, 28], although there is no universally agreed upon method; our choice is comparable to those in [7, 3, 16].

Straightforward analysis shows that the GS and SOR algorithms require 14 and 18 floating point operations (flops) per pixel per iteration respectively. The number of iterations required for convergence of the iterative algorithms grows with image size [21]. For reasonable size images (say, 512×512), SOR may require on the order of hundreds of iterations to converge, so that the total computation per pixel can be on the order of 10^3 to 10^4 flops. On the other hand, the MR algorithm requires 76 flops per pixel (see Appendix B). Note further that *the MR algorithm is* not *iterative.* Thus, as we will now see, the computational gain associated with the MR algorithm can be on the order of one to two orders of magnitude for problems of this size and substantially greater for problems defined over much larger spatial regions.

3.2 Rotation Sequence

We begin with a comparatively small synthetic example of rotational motion in order to illustrate the basic features of our approach. Specifically, this first example is a synthetic sequence of Gaussian images modulated by a spatial sinewave with the first frame brightness pattern given by:

$$E(z_1, z_2, t_1) = \sin(\text{atan}(z_1 - 23, z_2 - 28)) \exp(-\frac{1}{2} z' Z^{-1} z) \tag{60}$$

$$z = \begin{bmatrix} z_1 - 23 \\ z_2 - 28 \end{bmatrix} \tag{61}$$

$$Z = \begin{bmatrix} 1000 & 0 \\ 0 & 500 \end{bmatrix} \tag{62}$$

where $\text{atan}(z_1, z_2)$ is a 2π arctangent ($\text{atan}(0,1) = 0$, $\text{atan}(1,0) = -\pi$), $h = 1$ and $M = 6$ (i.e. the image lattice is 64×64, cf. the discussion about discretization at the beginning of

Section 2.1). The second frame is equal to the first, rotated by 1 degree about pixel (23,28). The first frame and actual optical flow are illustrated in Figure 4. The rms value of this flow field is 0.49.

The first point we wish to examine is the visual appearance of the estimates produced. Figure 5 shows four different estimates of the optical flow. The first of these (a) is the SC estimate produced using the original smoothness constraint formulation and performing 50 iterations of the SOR algorithm[6]; (b) is the finest scale of the MR estimates produced by the MR algorithm with the parameters set as $b = \mu = 1$; (c), which we refer to as MR-PF, is a post-filtered version of the MR estimates in (b) to be described; and (d), which we refer to as MR-SOR, is the estimate produced by performing 5 iterations of the SOR algorithm used in (a) but using the MR estimates in (b) as an initial condition. All four estimates clearly display the rotational nature of the true flow with quality that is roughly comparable. In particular, while rms error is not necessarily an appropriate measure of absolute estimate quality, it is of value in assessing the *relative* quality of these four methods, and for this example the rms errors for the estimates in Figure 5 are:

(SC)	0.24
(MR)	0.22
(MR-PF)	0.22
(MR-SOR)	0.20

which indicates that the MR method and its variations in (c) and (d) yield estimates of quantitative accuracy comparable to the SC-based method.

Despite this fact, the MR estimate in (b) has visual characteristics that may be somewhat distracting to the viewer: namely, the apparent blockiness of the estimates. As the rms errors indicate, and as we argue further in a moment, this visual artifact is not quantitatively significant. However, its nature and the reason for its presence motivate the computationally simple post-processing procedures illustrated in parts (c) and (d) of Figure 5. The first of these is motivated by the interpretation of our MR algorithm in terms of wavelet transforms and multiresolution analysis [6, 23]. Specifically, a natural interpretation of our model is that of providing multiresolution approximations of an image or random field; i.e. the values of a quadtree process at a given scale can be thought of as the so-called "scaling coefficients" [23] of particular basis functions used in the approximation at that scale. In that sense, the flow field estimate in (b) corresponds to the Haar approximation in which the basis functions are piecewise constant over squares of size corresponding to the scale being represented. The blockiness in (b) is thus due to the "staircase" nature of the Haar approximation. On the other hand, there are far smoother choices for basis functions and multiresolution approximations, each of which corresponds in essence to convolving the 2-D array of quadtree estimates at the finest scale with particular FIR filters. The MR-PF estimates in Figure 5(c) corresponds to using the FIR filter given by (59) together with the MR estimate in (b).

The estimate in (d) is motivated by the observation that the visual artifacts in the estimate (b) are local and high-frequency in nature. Indeed, it is *precisely* these high frequency artifacts that are quickly and easily removed by SOR or GS algorithms computing

[6]In this and subsequent examples, the iterative algorithms computing the solution of (4) were initialized with zero.

the smoothness constraint solution. This is clearly demonstrated in the MR-SOR estimates in (d) in which only 5 SOR iterations have been used to post-process (b).

Let us now turn to the question of computational complexity. Figure 6 illustrates the rms error in the flow estimates as a function of iteration for the SOR and GS algorithms. The rms error in the MR flow estimate of Figure 5(b) as well as those of MR-PF and MR-SOR in (c) and (d) are also indicated in the figure. The procedures used to generate the MR, MR-PF and MR-SOR estimates are *not* iterative and thus the associated rms errors are shown simply as straight lines. Note first that, as expected, the SOR algorithm is significantly faster than the GS algorithm (they will converge to the same result since they are solving the same partial differential equation). However, the SOR algorithm itself has a substantial computational burden. For example, while the SOR algorithm has not converged after 50 iterations, the estimates in Figure 5(a) are not bad, but even at this point and even for this small example, SOR requires far more computation than the MR based estimate. In particular, as we indicated previously, the computational load of the MR algorithm equals 4.2 SOR iterations, while producing the MR-PF and MR-SOR estimates requires computation equivalent to 5.6 and 9.2 SOR iterations, respectively[7]. Thus, for this small example, the algorithms corresponding to Figures 5(b) – (d) offer computational savings over SOR of factors of $50/4.2 = 11.9$, $50/5.6 = 8.9$ and $50/9.2 = 5.4$ respectively. As an aside, note that these results also suggest that if one insists upon solving the partial differential equation corresponding to the SC formulation, then using the MR estimate as an initial condition is a computationally attractive way in which to do this. Specifically, Figure 7 illustrates the rms difference between the smoothness constraint solution[8] and the intermediate values of the GS, SOR and MR-initialized SOR estimates as a function of iteration. The error plot for the MR-initialized SOR algorithm begins at 4.2 iterations to take into account the initial computation associated with the MR algorithm. The figure demonstrates that the MR-initialized SOR approach provides a substantial reduction in computational burden even for this small size problem. This in fact suggests that MR algorithms may be of more general use in the efficient solution of partial differential equations in other applications as well.

As we have emphasized, the MR algorithm has other attractive features beyond its computational efficiency, including the fact that it directly provides estimates at multiple resolutions. Figure 8 depicts these estimates at scales $m = 1, 2$ and 3 (where the finest scale $m = 6$ estimates are in Figure 5(b)). These coarser estimates also obviously capture the rotational motion and may, in some cases, be preferable representations of perceived motion because of their comparative parsimony compared to Figure 5(b). Indeed in many applications one is interested in fairly aggregate measures of motion which these estimates provide directly. Furthermore, as we describe next, the MR algorithm in fact directly provides a precise way in which to determine the optimal resolution for characterizing optical flow in different regions of the image, the basis of which is the multiscale covariance information computed by the MR algorithm.

[7]With respect to the MR-PF estimates (c), straightforward convolution of the two components of the optical flow in (b) with a separable 7×7 filter requires 26 flops per pixel (equivalent to 1.4 SOR iterations) and could, of course, be reduced further with FFT algorithms.

[8]The smoothness constraint solution is approximated as the SOR algorithm optical flow estimates after 500 iterations.

Figure 9 illustrates the trace of the 2×2 estimation error covariance in (52) at each point in the quadtree at different scales. Bright areas correspond to regions of lower covariance (higher confidence). Note that around the border of the image, where the Gaussian has tapered off and the gradients are relatively small, the error covariance is relatively large, as compared to the region around the point of rotation. One use of this covariance information is to provide information that may be useful to higher level vision algorithms which use the optical flow field in conjunction with information from other sources, and need to combine this information in a rational way. Moreover, as we have suggested, this information can also be used in the context of addressing the problem of resolution vs. accuracy in the estimates. The idea is that we would expect to estimate rather well the coarse resolution features in the optical flow field and that finer resolution features could be estimated with decreasing fidelity depending on the quality and characteristics of the available data (e.g. on the presence or absence of fine scale image intensity fluctuations). Thus, what we would like is a rational procedure for determining the estimate resolution supported by the data.

There are several ways in which the flow estimate covariance information can be used to approach this problem. One possibility, which has a precise statistical interpretation, is as follows. To each node at the finest scale, we can trace a path up to the root node, where nodes in the path correspond to the parent, grandparent, great-grandparent, etc. of the node at the finest level. The optical flow estimates at each of these resolutions can be thought of as successively coarser representations of the optical flow estimate at the finest scale. Associated with that same path is a sequence of smoothing error covariance matrices computed via (52). At each pixel location we can choose the optimal resolution at which to represent the field by choosing the scale at which this error covariance is minimum. In Figure 10 the scale of the minimum of the trace of the smoothed error covariance along this path is plotted for each lattice site. Note that in regions near the border, where the Gaussian has tapered off and not much gradient information is available, a lower resolution representation for the flow field is given. On the other hand, near the point of rotation, where there is gradient information, the resolution is at a higher (i.e. finer) level. It is interesting to note that the areas in which the finest level MR estimate of Figure 5(b) has the most visually obvious blocky behavior are also areas in which one has no business estimating optical flow at such a fine scale to begin with. Said another way, one interpretation of Figure 10 is that *any* estimate of optical flow at such a fine scale in such regions is a visual artifact!

Finally, let us briefly comment on the choice of the parameters b and μ in the MR algorithm. In particular, we have found through experimentation that the rms error in the estimates and their qualitative appearance is relatively insensitive to b and μ. Figure 11 depicts the rms errors in the MR flow estimates for the rotation example as a function of b and μ, displaying characteristically flat behavior over a very large range of values.

3.3 Yosemite Sequence

The second example is a synthetic image sequence which simulates the view from a plane flying through the Yosemite Valley[9]. The first image in the sequence and the corresponding

[9]This sequence was synthesized by Lyn Quam of SRI International. The original sequence is 252×312. As discussed in Appendix A, it is straightforward to adapt our approach to trees other than regular quadtrees, i.e. to trees with varying numbers of branches. However, for simplicity, in these experiments we have coded

optical flow are shown in Figure 12. The rms value of the flow field is 1.86.

Figure 13 illustrates four estimates of the optical flow corresponding to (a) the SC formulation after 100 iterations of the SOR algorithm, (b) the finest scale of estimates produced by the MR algorithm with parameters $b = \mu = 1$, (c) the MR-PF estimates derived as described previously and (d) the MR-SOR estimates produced by post-processing the MR estimates with 10 iterations of SOR. The estimates are qualitatively similar, each indicating the fly-through nature of the sequence. The estimates are also quantitatively similar as indicated by the rms errors for the four estimates:

(SC)	0.76
(MR)	0.79
(MR-PF)	0.79
(MR-SOR)	0.78

The rms errors as a function of iteration are shown in Figure 14. Note that the SC estimates (a) have actually not yet converged after 100 iterations and that when they do, the rms error of the SC estimate is slightly higher than those for the various approaches based on the MR algorithm.

Again, there is some blockiness in the MR optical flow estimates, and, as seen in Figures 13(c) and (d), some of this effect can be eliminated by post-processing the estimates with an FIR filter as in the previous example. There is still some blockiness apparent, but comparison with (a) shows that this is also apparent in the SC solution. Hence, the residual blockiness in the smoothed estimates is not due to the quadtree structure, but rather to the nature of the image sequence data itself.

An examination of computational complexity again shows the gains achievable using MR-based methods. The SC flow estimates shown in Figure 13(a) required 100 SOR iterations in this example, representing a factor of $100/4.2 = 23.8$ more computation than the MR estimates. Likewise, the MR-PF and MR-SOR (c) and (d) represent factors of $100/7.7 = 13$ and $100/14.2 = 7.0$ computational improvement. In general, the number of iterations required for convergence of the SOR algorithm for the SC formulation depends on several things, including the parameter R, the image gradient characteristics and the image size. Analysis in [21] shows that the SOR algorithm requires on the order of N iterations for an $N \times N$ image. Thus, we expect substantially more computational savings as the image size increases.

Furthermore, as before one would expect to be able to quickly obtain the SC solution by using the MR solution as an initial condition. Figure 15 illustrates how the GS, SOR and MR-initialized SOR algorithms converge to the smoothness constraint solution. Note that visually, there is almost no difference between the MR-initialized SOR estimates Figure 13(d) and the SC estimates shown in Figure 13(a). Indeed, the rms difference between the MR estimates and the smoothness constraint solution is 0.178, while the rms difference between the estimates in Figure 13(a) and the smoothness constraint solution is 0.181. More generally, Figure 15 shows that for any given number of iterations, the MR-initialized SOR estimates are substantially closer to the final solution than the GS or SOR estimates.

our algorithms for quadtrees. For this example, then, we extracted a 252×256 portion of the sequence (the left side) so that processing could be done on a quadtree with 256×256 lattice sites at the finest level. The measurement matrix $C(s)$ defined at the unneeded four rows of the quadtree structure was set to zero, reflecting the fact that we have no information about the (non-existent) optical flow field in that region.

Estimates of the optical flow at scales $m = 1, 2, 3$ computed via the MR algorithm are shown in Figure 16 and multiscale error covariance images, again, corresponding to the traces of the smoothing error covariance matrices at individual lattice sites, are shown in Figure 17. The coarser versions of the flow are intuitively reasonable given the estimates at the finest level and, as expected, the covariance images are relatively dark (high covariance) in the top portion of the image where there is no gradient information available.

Figure 18 depicts a map of the optimum resolution for flow estimation at each pixel location computed as the minimum of the trace of the smoothed error covariance matrix along paths from nodes at the finest level to the root node. We see, not surprisingly, that the level of resolution chosen for the region with no intensity information is quite low. In addition, the resolution along the face of the mountain in the foreground is slightly reduced due to the relative lack of gradient information in the direction of the striations.

Finally, Figure 19 illustrates the variations in the rms error in the optical flow estimates to variations in the parameters b and μ. The figure shows that the estimates are relatively insensitive to the parameter b, and are also insensitive to μ for values ranging from slightly less than 1 upward. The degradation in performance as μ decreases toward zero is not uncommon or unexpected. In particular, as discussed in [11, 9, 10, 12, 52] decreasing μ leads to significant decreases in spatial correlation in the model and to far noisier sample paths. Thus, the estimates for small values of μ correspond to imposing virtually *no* smoothness constraint, resulting in estimated fields with noise-like characteristics. On the other hand, choosing any value of $\mu \geq 1$ yields results of comparable quality to each other and to the SC solution.

3.4 Moving Vehicle Sequence

The third example is based on a real[10] image sequence which depicts the view from a car driving down a road. The first image in the sequence is illustrated in Figure 20 and Figure 21 illustrates four estimates of the optical flow corresponding to (a) the SC formulation and 200 iterations of the SOR algorithm, (b) the finest scale of estimates produced by the MR algorithm with parameters $b = \mu = 1$, (c) the MR-PF estimate and (d) the MR-SOR estimate produced by post-processing the MR estimates (b) with 30 iterations of SOR.

Since the true optical flow is not available (as it was in the previous simulated examples), an alternate performance metric is needed. In particular, we will use a reconstruction error metric, which is often used in contexts in which one is interested in using optical flow for motion-compensated coding. This metric measures the mean square difference between the current image in a sequence and an estimate of it based on the computed optical flow, the previous image, and a bilinear interpolation scheme [29]. The optical flow used is that associated with the current image. Essentially, one estimates the brightness at any given point by using the optical flow to project that point back to the previous image. In general, that point will not be on the image plane, and the bilinear interpolation is required.

Figure 22 provides a comparison of reconstruction error performance for the approaches as a function of iteration (where once again the results for the non-iterative MR, MR-PF and MR-SOR approaches are depicted as horizontal lines). In this example, the SC solution was slightly better than the MR and MR-PF methods, achieving a slightly greater rms error

[10]The sequence is courtesy of Saab-Scania.

reduction from the value obtained without motion compensation (i.e. straightforward frame difference given by the zero-iteration starting point for SOR). However, this slight increase in performance is achieved at the cost of significantly greater computation. In particular, the computational gains are $200/4.2 = 47.6$, $200/6.98 = 28.7$ for the MR-PF and MR-SOR approaches, respectively. Furthermore, as is also illustrated in Figure 22, the modest performance gain of SC over MR can recouped with far less computation using the MR-SOR procedure which has a factor of $200/34.2 = 5.8$ computational speedup. Indeed, as Figure 23 shows, the MR-SOR solution of Figure 21(d) is *closer* to the SC solution than the result in Figure 21(a), which required 200 iterations of SOR to obtain.

As in the previous examples, multiresolution flow estimates and error covariance information is available at all levels of the quadtree, and an image of the error covariance information at the finest level lattice points is shown in Figure 24(a). Note in this case that the error covariance is relatively high (dark regions in the image) along the road where the image gradient is relatively low. Also, Figure 24(b) depicts the optimal resolution at which to recover the optical flow field computed using this error covariance information.

Finally, the sensitivity of the optical flow estimates in this example to parameter choice is shown in Figure 25. The figure shows that the reconstruction error is stable for $\mu \geq 1$ as in the Yosemite example, and is insensitive to variations in b over a significant range of values.

3.5 Chopper Sequence

The first frame of the real "chopper" sequence[11] is shown in Figure 26. Figure 27 illustrates four estimates of the optical flow corresponding to (a) the SC formulation and 200 iterations of the SOR algorithm, (b) the finest scale of estimates produced by the MR algorithm with parameters $b = \mu = 1$, (c) the MR-PF estimate and (d) the MR-SOR estimate produced by post-processing the MR estimates (b) with 80 iterations of SOR.

As in the previous example, rms reconstruction error is the metric we use for comparison since the true flow is not known. Figure 28 provides a comparison of the reconstruction error performance of the approaches as a function of iteration. Note that in this example all four methods yield essentially identical rms performance, but once again the MR-based algorithms have significant computational advantage. Computational gains for the MR, MR-PF, and MR-SOR approaches are $200/4.2 = 47.6$, $200/6.53 = 30.6$ and $200/84.2 = 2.38$.

Also, as in the previous examples, the performance of the MR algorithm is stable over a wide range of values of the parameters b and μ, as is illustrated in Figure 29. In addition, multiresolution estimates and error covariance information are, of course, available. For the sake of brevity, we illustrate only map of the optimum resolution information constructed from the multiscale error covariance information in Figure 30. Note in this case that the resolution level is relatively uniform over the image and is at a scale far coarser than the finest scale level (level 10). That is, the image spatial intensity variations in this image sequence are not particularly strong so that fine resolution flow estimation can only be achieved with high levels of uncertainty.

[11] The 480×480 image lattice was centered on the finest level of a 10 level (512×512 at the finest scale) quadtree. Again, as discussed in Appendix A, adapting our approach to deal directly with arbitrary size lattices is straightforward.

On the other hand, there is an important fine-level velocity feature of some significance in this image sequence, namely a helicopter, located near the center of the image frame, which is moving relative to the background. While the local image contrast in the image is not sufficiently strong to allow very accurate estimation of what is in essence a discontinuity in the optical flow field, it is reasonable to expect that there would be *some* useful, quantitative information in the image sequence that could be used to detect this motion discontinuity and obtain rough (i.e. coarse level) motion estimates. While it is beyond the scope of this paper to develop such a scheme in detail, we can provide an indication of how the MR method provides the essential elements for an effective solution.

The starting point for this is the well-known criterion of global smoothness constraint type formulations such as ours, namely that they tend to obscure localized motions such as that due to the helicopter in Figure 26. This is not surprising since SC-type formulations yield what are in essence low-pass spatial filters. However, there is an extremely critical point that is well-known in Kalman filtering theory and in that relating to the use of such filters for the detection of abrupt changes in time series or dynamic systems. Specifically, such filters can *also* be used to implement ***high-pass*** filters which produce outputs that not only *enhance* the discontinuities to be detected but also make optimal detection possible. Specifically, the residuals or innovations in a Kalman filter, that is, the difference between the observations and predicted observations based on model and data, represent a statistically whitened version of the observations resulting from what is in essence a high-pass filter. As discussed in many papers and books ([4, 51], for example), discontinuities in the data being processed then lead to distinctive ***signatures*** which can be looked for using optimal detection methods.

In a similar fashion we can compute the residuals of the MR estimates:

$$\nu(s) \quad = \quad y(s) - C(s)\widehat{x}^{s}(s) \tag{63}$$

for the chopper sequence, an image of which is illustrated in Figure 31. Note that in contrast to the original image in Figure 26, this residual image does not display *any* coherent structure other than the helicopter, making detection of the helicopter a far easier task in this domain. Furthermore, high pass filtering has in fact enhanced the chopper signature, as the helicopter rotors, nearly imperceptible in Figure 26 are clearly in evidence in Figure 31 because of the motion discontinuity. As we have indicated, statistically optimal methods for using residuals analogous to these have been developed for time series, and, as discussed in [4, 51], such methods require error covariance information from the estimator in order to specify the optimal detection procedure. Since the MR algorithm also produces such error covariance information it is possible to develop optimal detection methods in this imaging context as well. Such a method is currently under development.

4 Conclusions

We have presented a new approach to the regularization of ill-posed inverse problems, and have demonstrated its potential through its application to the problem of computing optical flow. This approach starts from the "fractal prior" interpretation of the smoothness constraint introduced by Horn and Schunck to motivate regularization based on a recently

introduced class of multiscale stochastic models. This new formulation leads to an extremely efficient, non-iterative, scale-recursive solution, yielding substantial savings over the iterative algorithms required for the smoothness constraint solution. In particular for 256×256 or 512×512 images, our algorithm leads to computational savings on the order of a factor of 10 to 100. Indeed, since the iterative approaches associated with the smoothness constraint solution typically take longer to converge as the image grows, whereas the per pixel computation associated with the MR algorithm is independent of image size, even larger savings can be realized for larger image domains.

Our approach has a number of potential advantages in addition to the reduction in computational cost. First, multiresolution estimates of the flow field are available and, although we have not taken advantage of it in this paper, the MR algorithm also allows for *multiresolution measurements* of the optical flow, i.e. measurements as in (25) but for triples $s = (m, i, j)$ at *several* scales. Second, error covariance information is available, allowing one to assess the quality of the estimated optical flow, and we have used this information to suggest one means of addressing the resolution vs. accuracy tradeoff inherent in ill-posed problems by specifying the optimal resolution for flow reconstruction at each point in the image. Finally, the MR algorithm provides an excellent initialization for algorithms computing a solution based on a smoothness constraint formulation.

While we have not pursued it here, the multiresolution philosophy introduced here may offer a promising approach to motion-compensated image sequence coding. In particular, although we used the coding metric of reconstruction error as the basis for the comparison of the SC and MR approaches, the methods presented here would not be the method of choice in a coding context. In particular, motion-compensated coding algorithms designed specifically to minimize this criterion [2, 29, 50] will generally outperform the SC and MR approaches (which are not designed for that express purpose). However, the computationally efficient MR algorithm can be used as an initial preconditioning step for such coding algorithms. In addition, one can also imagine a second way in which MR ideas could be used in this context. In particular, one of the problems with the SC and MR based methods is the differential form of the brightness constraint which, given the discrete nature of spatial and temporal sampling, is only valid for relatively small interframe displacements. In contrast, methods such as [2, 29, 50] use a direct displaced frame matching metric, which is nothing but the integrated version of the brightness constraint. A common approach to dealing with larger displacements with the differential brightness constraint is to spatially blur the image sequences, i.e. to consider lower resolution versions of the image to estimate larger displacements [14, 18]. What this suggests is an MR approach in which we not only have a multiresolution model for optical flow but also multiresolution measurements. The development of such an approach remains for the future.

Also, the framework in which our method is developed suggests a method for directly detecting unmodeled discontinuities in the optical flow field in a rational and statistically optimal way. In particular, the measurement residual field represents a high-pass version of the observed data which accentuates the effects of motion discontinuities and removes other features corresponding to smoothly varying parts of the flow field. For time series, such residuals provide the basis for extremely effective methods for the detection of discontinuities, and the development of corresponding methods in our multiscale, image processing framework represents a promising direction for the future. Indeed, this suggests a number

of additional directions for extending time-series methods to the imaging context such as adaptive estimation of the multiscale parameters b and μ in order to adaptively adjust the level and nature of the regularization imposed on different image regions. While such adaptive methods are certainly not unknown in image processing, our scale-recursive framework not only leads to an extremely efficient framework for the realization and provides the error covariance information needed for the development of statistically optimal methods but the use of a pyramidal framework provides enormous flexibility in adaptation. For example, in the time series case, the use of a very large value for the noise parameter corresponding to b at some point in time essentially decouples the processing before and after that point (since no smoothness at that point is expected). In our framework a large value for b at some node decouples the processing within the region, corresponding to the subtree of pixels beneath that node, from processing outside that region, exactly what would be needed to deal with a region corresponding to motion discontinuity relative to the background.

Finally, in this paper we have focused on a particular image processing problem, the computation of optical flow. However, we believe that the multiscale stochastic modeling approach can be more generally useful. In particular, it may provide a computationally attractive alternative to standard approaches to the broad class of estimation problems in which the underlying field to be estimated is modeled as a Gaussian Markov random field or as the solution of noise driven partial differential equations, or in which a "smoothness constraint" type regularization is employed. Viewing the multiscale models as an alternative underlying model should lead to significant computational savings for such problems and should also have the other benefits we have described.

Acknowledgment: *The authors gratefully acknowledge the contributions of Dr. Albert Benveniste who, among other things, first suggested that the problem of computing optical flow would be well suited to this approach. We are also pleased to acknowledge the reviewers for their comments that have helped to enhance the development and presentation of these results.*

A Non-homogeneous Tree Structures

We made the assumption at the beginning of Section 2 that the image lattice is square, and that the number of rows is equal to a power of two. The reason we have done this is because of the fact that the multiscale model described in this paper is defined on a quadtree structure. There are at least two ways to relax the assumption. First, we could simply zero pad $C(s)$ on the image lattice to make it fit the quadtree structure. This corresponds assuming no information is available about the (non-existent) optical flow in that region. A second, slightly more elegant approach, would be to change the modeling structure to accommodate the lattice. In particular, we would like to have a structure which gives us the proper number of nodes on the finest level. The quadtree structure is homogeneous in the sense that each parent has four offspring; what we are proposing are non-homogeneous tree structures in which different parents may have different numbers of offspring. For example, suppose one had a 6×9 lattice. Figure 32 illustrates a sequence of grids that one might use to model a random field defined this lattice. In the first level, the root node has six offspring, two in the row direction and three in the column direction. At the second level, each node has nine offspring, three in the row direction and three in the column direction. Thus, at the finest level there is a 6×9 lattice. This example illustrates only one simple suggestion. More complicated tree structures could be derived, and certainly the idea could be combined with zero padding.

B MR Algorithm Complexity Analysis

In this section we analyze the computational complexity of the MR algorithm. The analysis applies to the specific model given by (54) - (58). The model is repeated here for convenience:

$$
\begin{aligned}
x(s) &= x(s\tilde{\gamma}) + (b4^{\frac{-\mu m(s)}{2}})w(s) && (64)\\
y(s) &= C(s)x(s) + v(s) && (65)\\
w(s) &\sim \mathcal{N}(0, I) && (66)\\
v(s) &\sim \mathcal{N}(0, R(s)) && (67)\\
x_0 &\sim \mathcal{N}(0, pI) && (68)
\end{aligned}
$$

where $R(s) = \max(||C(s)||^2, 10)$. The analysis below takes into account all floating point adds, multiplies and divides.

Consider first the update step given by (40) - (43). $P(s|s+)$ is initialized with pI. Computation of $V^{-1}(s)$ requires 6 floating point operations (the inverse requires 1 divide since $V(s)$ is a scalar and the comparison required to compute $R(s)$ is not counted). Computation of $K(s)$ requires 3 flops. Computation of $P(s|s)$ requires 7 flops (Perform the $C(s)P(s|s+)$ first, and use the fact that $P(s|s)$ must be symmetric). Initialize $\widehat{x}(s|s+)$ with zero. Computation of $\widehat{x}(s|s)$ then requires 2 flops. The update step is required only at the finest level, since this is the only place we have data for in the optical flow problem. Thus, the total computation associated with this step is 18×4^l flops (l is defined to be the number of levels in the quadtree. There are 4^l points at the finest level.)

Next, consider the prediction step, (44) – (46). Computation of $Q(s\alpha_i)$ is negligible because this parameter varies only as a function of scale (level). Computation of $P(s|s\alpha_i)$ requires 5 flops (note that $F(s)$ and $Q(s\alpha_i)$ are diagonal multiples of the identity). Computation of the predicted estimate $\hat{x}(s|s\alpha_i)$ requires 2 flops. These computations must be done at levels 1 through l. Thus, the total computation associated with this step is approximately $7 \times 4/3 \times 4^l$ flops.

Next, consider the merge step, (47) – (48). Computation of $P(s|s+)$ requires 44 flops (there are five 2×2 inverses requiring 6 flops apiece, and the computation of $(1-q)P_s^{-1}$ is negligible since it only varies with scale. The inverses require only 6 flops because the matrices involved are 2×2 and symmetric.) Computation of $\hat{x}(s|s+)$ requires 36 flops. The merge step must be done at levels 0 through $l-1$. Thus, the total computation associated with this step is $80 \times 1/3 \times 4^l$ flops.

Finally, consider the steps in the downward sweep, (51) – (53). Computation of $J(s)$ requires 12 flops (the matrix $P(s\bar{\gamma}|s)$ has already been inverted in (48), $F(s)$ is a multiple of the identity and $J(s)$ is symmetric.) Computation of $P^s(s)$ is not required, unless one is explicitly interested in the error covariance of the smoothed estimate. Computation of $\hat{x}^s(s)$ requires 10 flops. The smoothing step must be done at levels 1 through l. Thus, the total computation associated with this step is 22×4^l flops.

We can now add up all of the computations associated with the MR algorithm. There are 4^l pixels in the problem domain, and thus the algorithm requires $18 + 28/3 + 80/3 + 22 = 76$ flops per pixel. We note that this is a lower bound on the number of flops per pixel in any implementation of the algorithm and that the implementation with the lowest number of flops per pixel may not be the best. The reason is simply that there may not be enough memory available to keep all intermediate calculations around (such as the inverses computed in (48) and reused in (53)). We compute the complexity of the GS and SOR algorithms in the same way (i.e. all intermediate results are assumed to be available), and thus the computational comparison we make between these algorithms is based on optimal (in terms of the number of flops) implementations. Suboptimal implementation of the MR algorithm will lower its computational advantage, but any reasonable implementation (for instance one which saves just $\hat{x}(s|s)$, $P(s|s)$ and the measurement data) will still provide a significant savings over the SOR and GS algorithms.

References

[1] J. Aggarwal and N. Nandhakumar. On the computation of motion from sequences of images – a review. *Proceedings of the IEEE*, 76:917–935, 1988.

[2] N. Baaziz and C. Labit. Multigrid motion estimation on pyramidal representations for image sequence coding. Technical Report 572, IRISA, February 1991.

[3] J. Barron, D. Fleet, S. Beauchemin, and T. Burkitt. Performance of optical flow techniques. In *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, pages 236–242, 1992.

[4] M. Basseville and A. Benveniste. *Detection of Abrupt Changes in Signals and Dynamical Systems*, volume 77 of *Lecture Notes in Control and Information Sciences.* Springer-Verlag, 1986.

[5] M. Bertero, T. Poggio, and V. Torre. Ill posed problems in early vision. *Proceedings of the IEEE*, 76:869–889, 1988.

[6] P. Burt and E. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31:482–540, 1983.

[7] T. Chin. *Dynamic Estimation in Computational Vision.* PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 1991.

[8] T. M. Chin, W. Karl, and A. Willsky. Sequential optical flow estimation using temporal coherence. Technical Report LIDS-P-2122, Massachusetts Institute of Technology Laboratory for Information and Decision Systems, 1992. Submitted to IEEE Transactions on Image Processing.

[9] K. C. Chou. *A stochastic modeling approach to multiscale signal processing.* PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, May 1991.

[10] K. C. Chou, A. S. Willsky, and A. Benveniste. Multiscale recursive estimation, data fusion and regularization. Technical Report LIDS-P-2085, Massachusetts Institute of Technology Laboratory for Information and Decision Systems, 1991.

[11] K. C. Chou, A. S. Willsky, A. Benveniste, and M. Basseville. Recursive and iterative estimation algorithms for multiresolution stochastic processes. In *Proceedings of the IEEE CDC*, 1989.

[12] K. C. Chou, A. S. Willsky, and R. Nikoukhah. Multiscale systems, Kalman filters and Riccati equations. Technical report, Massachusetts Institute of Technology Laboratory for Information and Decision Systems, 1992.

[13] S. C. Clippingdale and R. G. Wilson. Least squares image estimation on a multiresolution pyramid. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, 1989.

[14] W. Enkelmann. Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences. *Internation Journal of Computer Vision*, 43:150–177, 1990.

[15] D. Fleet and A. Jepson. Computation of component image velocity from local phase information. *Internation Journal of Computer Vision*, 5:77–104, 1990.

[16] N. Haddadi and C. C. Jay Kuo. Computation of dense optical flow fields with a parametric smoothness model. Technical Report USC SIPI 233, USC Signal and Image Processing Institute, 1992.

[17] D. Heeger and E. Simoncelli. Sequential motion analysis. In *Proceedings of the AAAI Robot Navigation Symposium*, March 1989.

[18] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using Markov random fields. Technical Report 561, Programme 6, IRISA, January 1991.

[19] B. K. P. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[20] J. Konrad and E. Dubois. Estimation of image motion fields: Bayesian formulation and stochastic solution. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1072–1074, 1988.

[21] C.-C. J. Kuo, B. C. Levy, and B. R. Musicus. A local relaxation method for solving elliptic pde's on mesh connected arrays. *SIAM J. Sci. Stat. Comput*, 8:550–573, 1987.

[22] M. Laurentev, V. Romanov, and S. Shishat-skii. *Ill-posed problems of mathematical physics and analysis*. American Mathematical Society, 1986.

[23] S. Mallat. Multi-frequency channel decomposition of images and wavelet models. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37:2091–2110, 1989.

[24] B. Mandelbrot and H. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10:422–436, October 1968.

[25] A. Mariano. Contour analysis: A new approach for melding geophysical fields. *Journal of Atmospheric and Oceanic Technology*, 7:287–295, 1990.

[26] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. Technical Report CICS-P-88, Center for Intelligent Control Systems, 1988.

[27] D. Murray and B. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:220–228, 1987.

[28] H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593, 1986.

[29] A. Netravali and J. Robbins. Motion-compensated television coding: Part 1. *Bell System Technical Journal*, 58:631 – 670, 1979.

[30] R. Pool. Making 3-D movies of the heart. *Science*, 251:28–30, 1991.

[31] P. Prenter. *Splines and Variational Methods.* Wiley, 1975.

[32] J. Prince and E. McVeigh. Motion estimation from tagged MR image sequences. *IEEE Transactions on Medical Imaging*, 11:238–249, 1992.

[33] D. Raviv. A quantitative approach to camera fixation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Maui, Hawaii, 1991.

[34] A. Rougée, B. Levy, and A. S. Willsky. An estimation-based approach to the reconstruction of optical flow. Technical Report LIDS-P-1663, Massachusetts Institute of Technology Laboratory for Information and Decision Systems, 1987.

[35] A. Rougée, B. Levy, and A. S. Willsky. Reconstruction of two-dimensional velocity fields as a linear estimation problem. In *Proceedings of the 1st International Conference on Computer Vision*, pages 646–650, 1987.

[36] J. Schroter and C. Wunsch. Solution of nonlinear finite difference ocean models by optimization methods with sensitivity and observational strategy analysis. *Journal of Physical Oceanography*, 16:1855–1874, 1986.

[37] A. Shio and J. Sklansky. Segmentation of people in motion. In *Proceedings of the IEEE Workshop on Visual Motion*, pages 325 – 332, Princeton NJ, 1991.

[38] E. Simoncelli, E. Adelson, and D. Heeger. Probability distributions of optical flow. In *IEEE Conference on Computer Vision and Pattern Recognition*, Maui, Hawaii, June 1991.

[39] A. Singh. Incremental estimation of image flow using a Kalman filter. In *Proceedings of the IEEE Workshop on Visual Motion*, pages 36–43, 1991.

[40] G. Strang. *Introduction to Applied Mathematics.* Wellesley-Cambridge Press, Wellesley MA, 1986.

[41] G. Strang. *Linear Algebra and its Applications.* Harcourt-Brace-Jovanovich, 1988.

[42] J. Stuller and G. Krishnamurthy. Kalman filter formulation of low-level television motion estimation. *Computer Vision, Graphics and Image Processing*, 21:169–204, 1983.

[43] V. Sundareswaren. Egomotion from global flow field data. In *Proceedings of the IEEE Workshop on Visual Motion*, pages 140–145, Princeton NJ, 1991.

[44] R. Szeliski. *Bayesian Modeling of Uncertainty in Low-level Vision.* Kluwer Academic, 1989.

[45] D. Terzopoulos. Image analysis using multigrid relaxation methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:129–139, 1986.

[46] A. Tikhonov and V. Arsenin. *Solutions of ill-posed problems.* Halstead Press, John Wiley and Sons, 1977.

[47] S. Turkoz. *Variational Procedures for ϕ^4 scalar field theory.* PhD thesis, Massachusetts Institute of Technology, Department of Physics, 1990.

[48] H. L. VanTrees. *Detection, Estimation and Modulation Theory: Part 1.* Wiley, New York, NY, 1968.

[49] J. Vlontzos and D. Geiger. A MRF approach to optical flow estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 853–856, 1992.

[50] D. Walker and K. Rao. Improved pel-recursive motion compensation. *IEEE Transactions on Communications*, 32:1128–1134, 1984.

[51] A. S. Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, pages 601–611, November 1976.

[52] G. Wornell. A Karhunen-Loeve like expansion for $1/f$ processes. *IEEE Transactions on Information Theory*, 36:859–861, 1990.

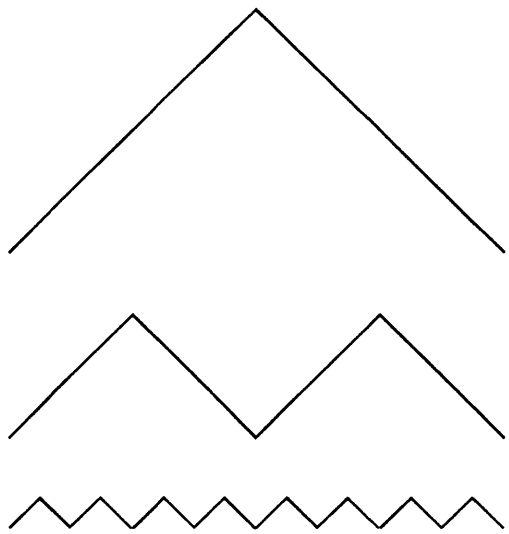


Figure 1: Depiction of three fields which are equally favored by the smoothness constraint, illustrating how this penalty provides a fractal prior model for the optical flow.

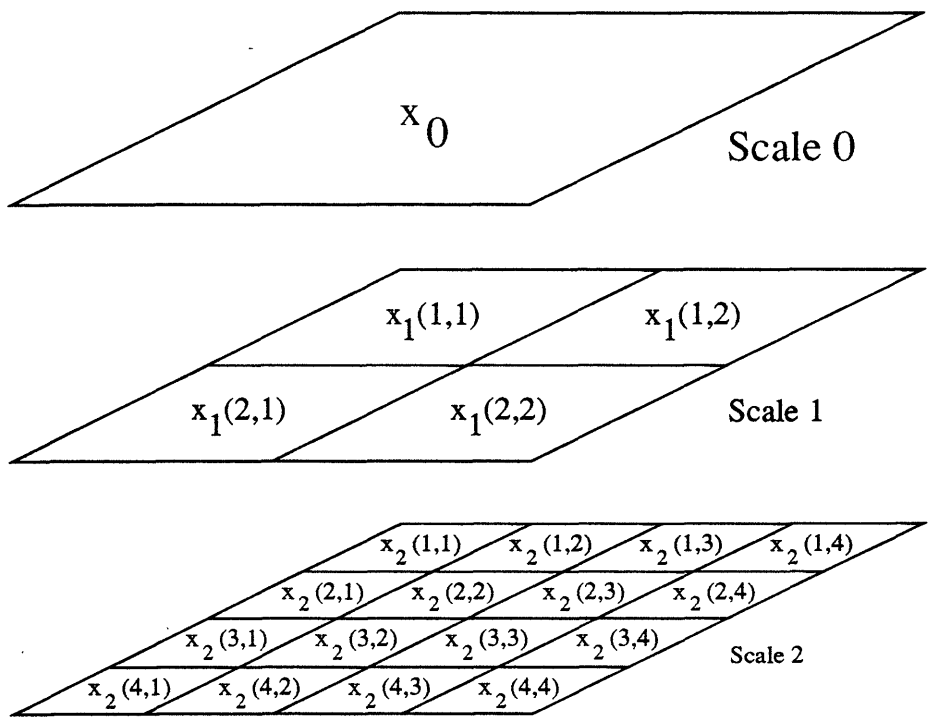


Figure 2: The structure of a multiscale optical flow field is depicted. The components of the field are denoted $x_m(i,j)$ where m refers to the scale and the pair (i,j) denotes a particular grid location at a given scale. At the coarsest scale, there is a single flow vector and, more generally, at the m^{th} scale there are 4^m vectors.

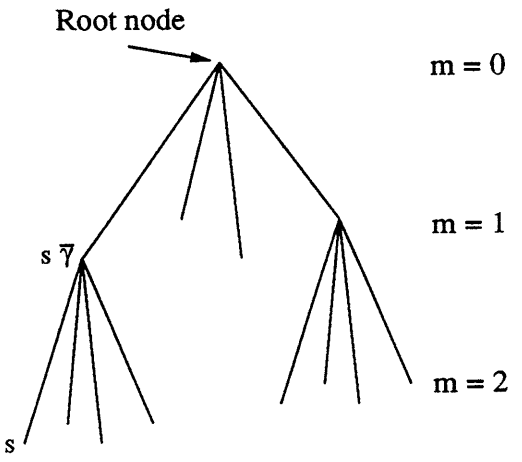
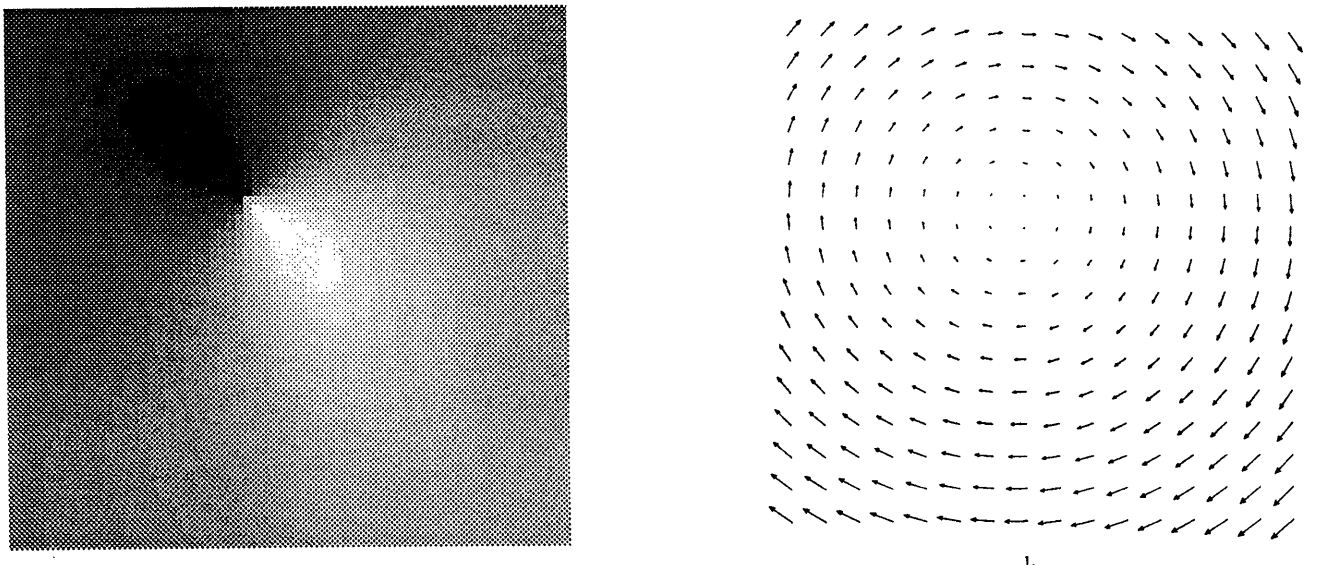
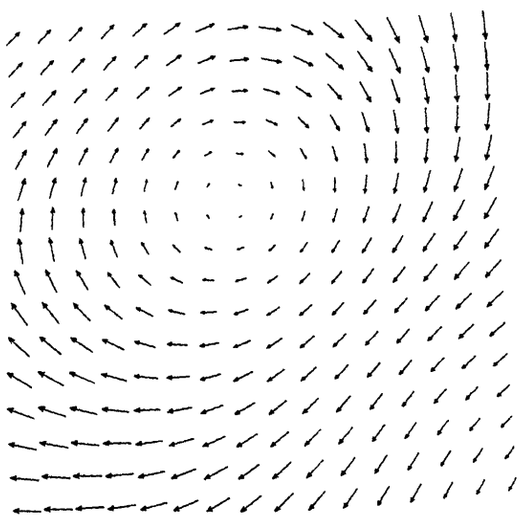


Figure 3: Quadtree structure on which the multiscale processes are defined. The abstract index s refers to a node in the quadtree; $s\bar{\gamma}$ refers to the parent of node s.

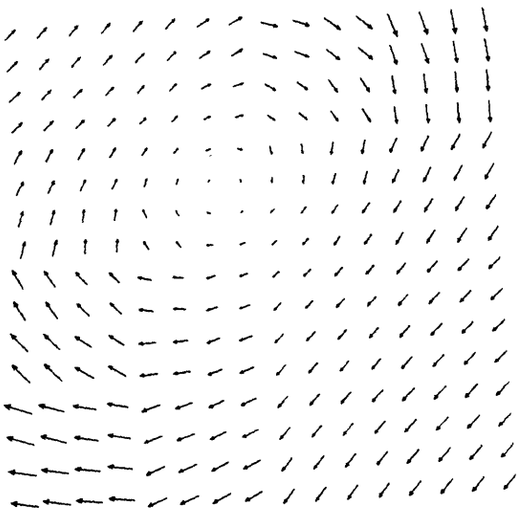


a b

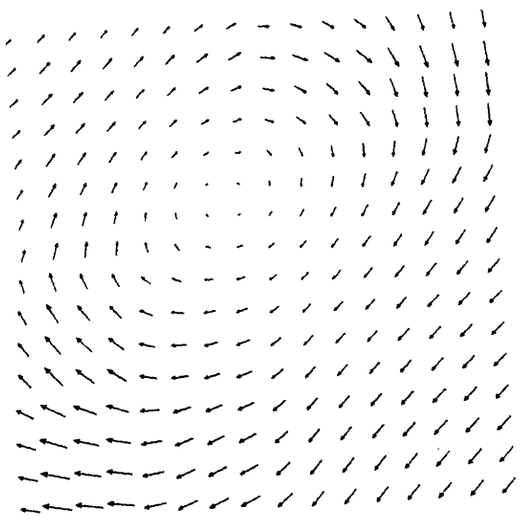
Figure 4: (a) First frame of the "rotation" sequence and (b) Rotation sequence true optical flow.



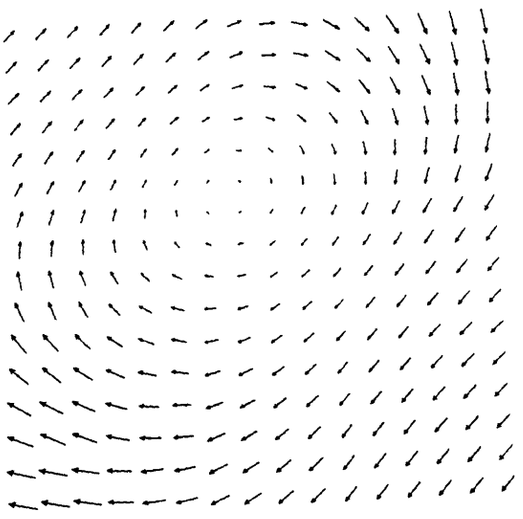
a



b



c



d

Figure 5: Rotation sequence flow estimates. (a) Smoothness constraint estimates computed using 50 iterations of SOR. (b) Multiscale Regularization (MR) estimates. (c) Post-filtered MR estimates and (d) Estimates produced by using MR estimates as initial condition for SOR algorithm.

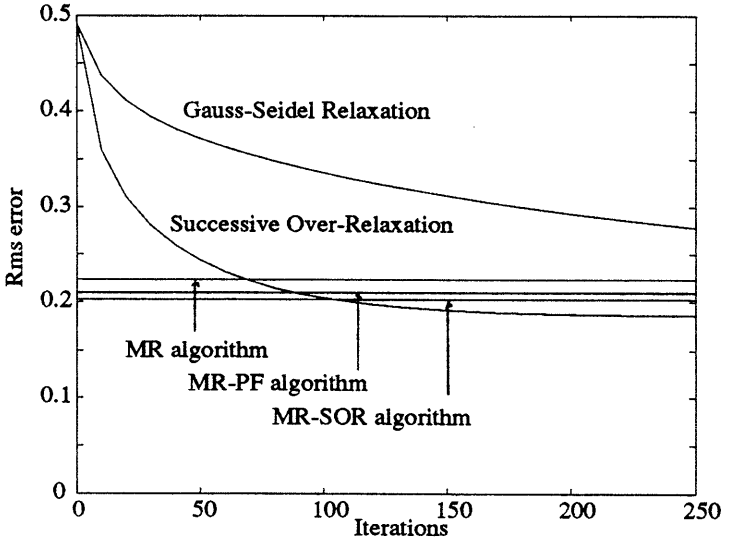


Figure 6: Rms Error Comparison of MR, MR-PF, MR-SOR, SOR and Gauss-Seidel (GS) algorithm flow estimates for the rotation sequence.

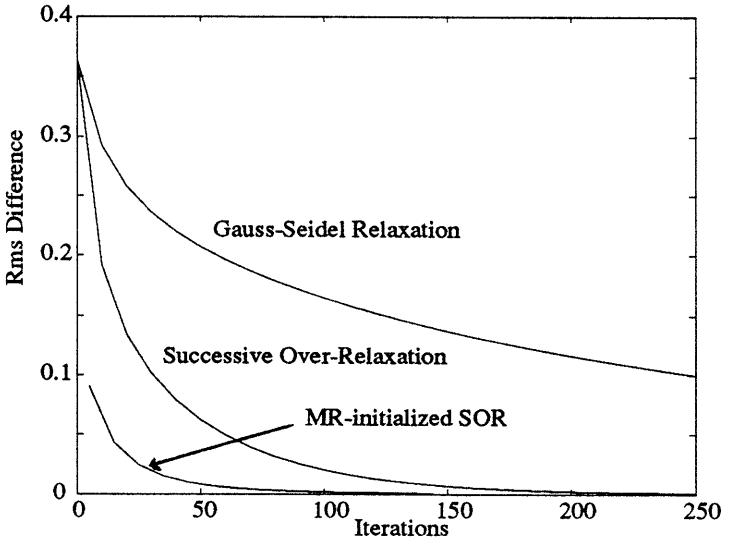
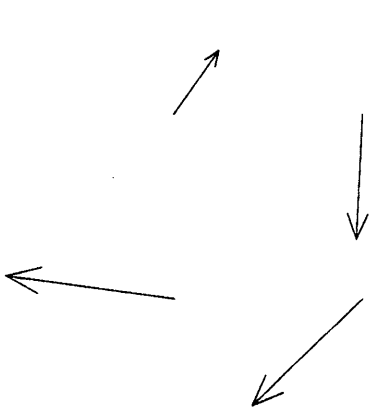
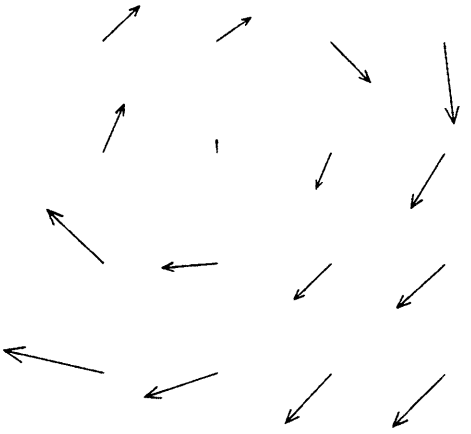


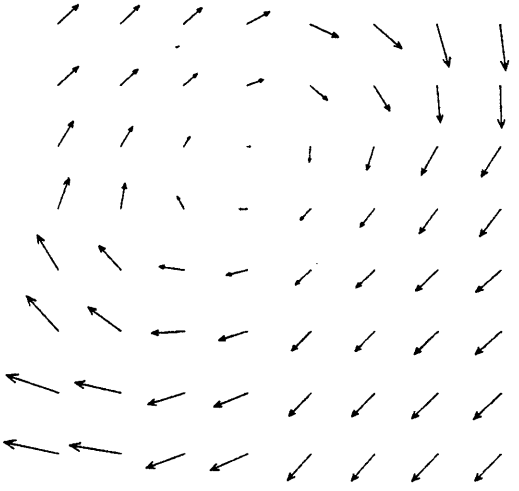
Figure 7: Rms difference comparison illustrates how the MR-initialized SOR, SOR and GS algorithms converge to the smoothness constraint solution for the Rotation sequence. The plots show the rms difference between the smoothness constraint solution and the estimates as a function of iteration. All will eventually converge, but the MR-initialized SOR algorithm converges much faster than SOR or GS.



a

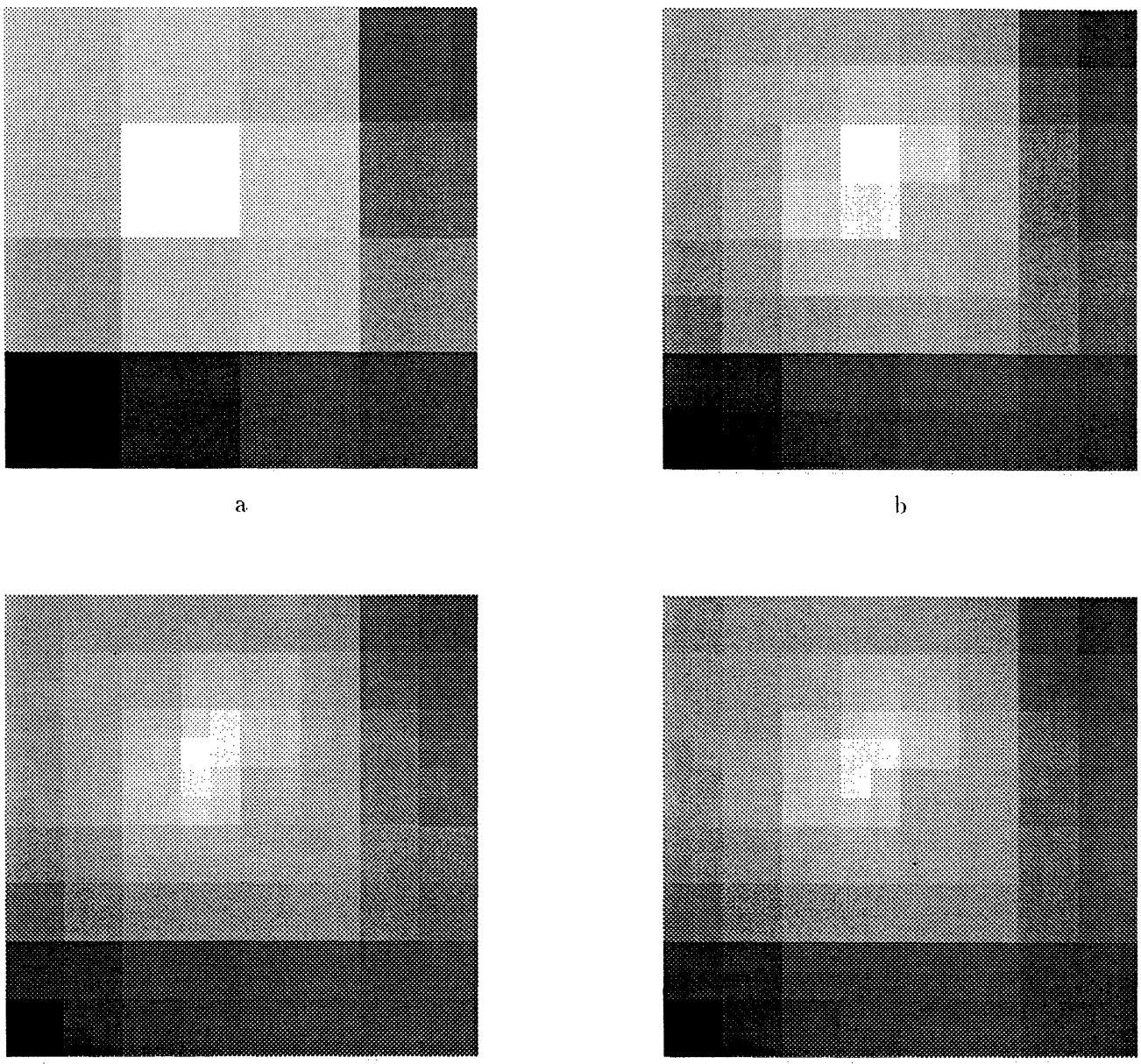


b



c

Figure 8: Multiscale Regularization flow estimates at the (a) first. (b) second and (c) third scales.



a b

c d

Figure 9: Multiscale Regularization error covariance at the (a) second, (b) third, (c) fourth and (d) sixth scales.

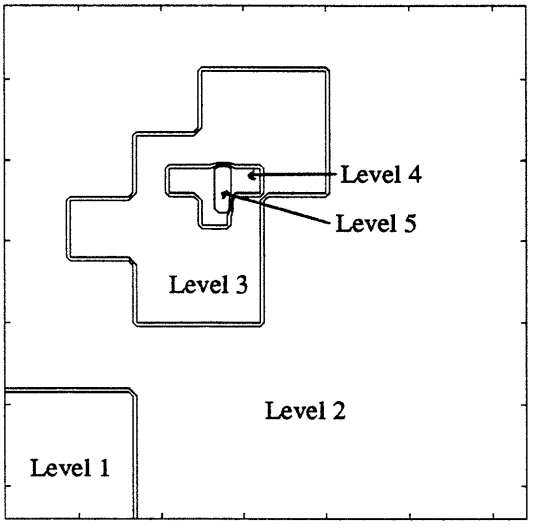


Figure 10: Map showing the optimal resolution for optical flow reconstruction for the rotation image sequence optical flow. At points near the focus of rotation the flow is represented at fine scales, while at points near the edge of the image (where little gradient information is available) the optical flow is represented at a coarser level of the quadtree.

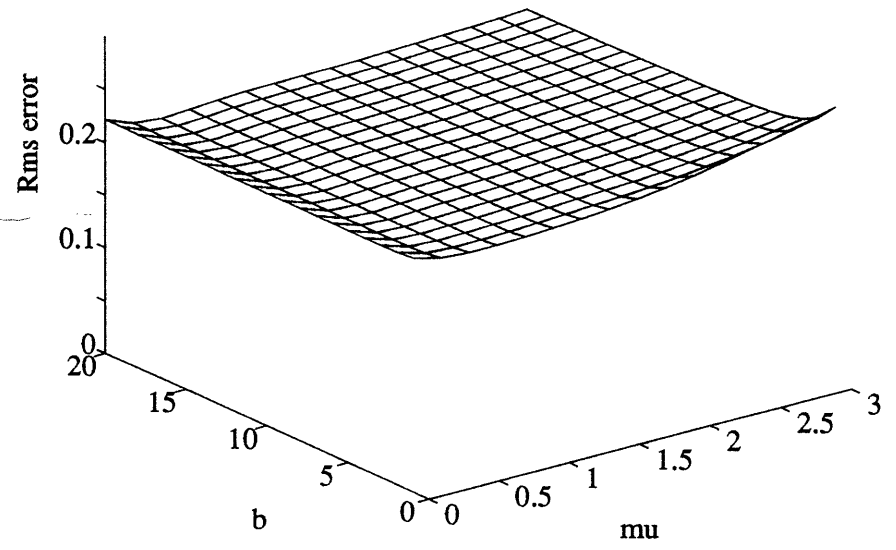
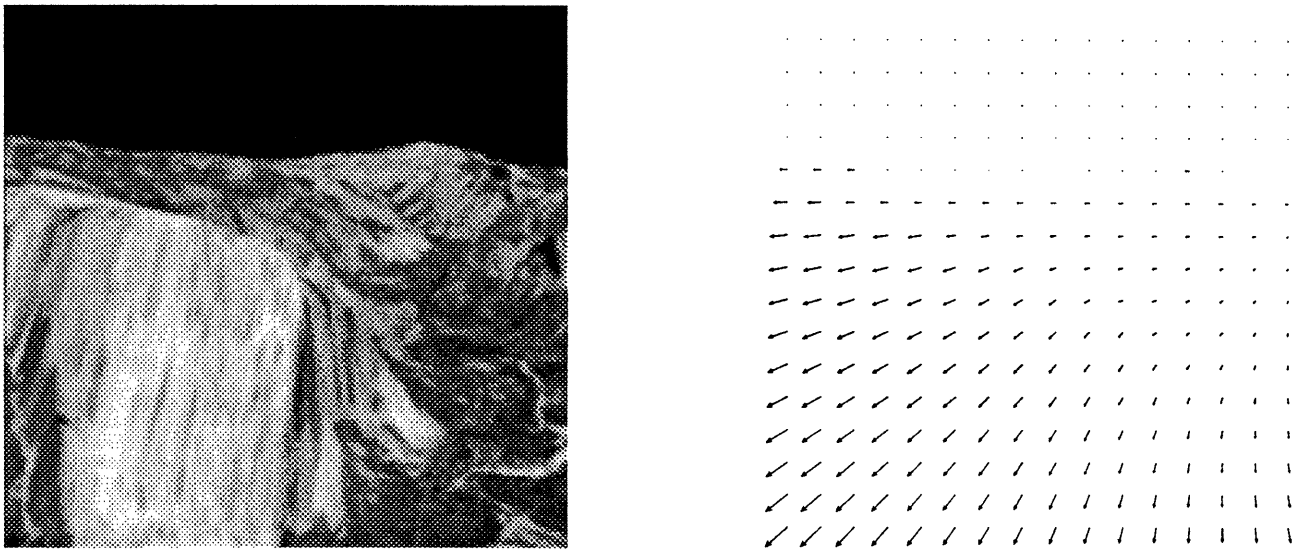


Figure 11: Multiscale Regularization rms error sensitivity to the parameters b and μ (rotation sequence).



a b

Figure 12: (a) First frame of Yosemite sequence and (b) Yosemite sequence true optical flow.

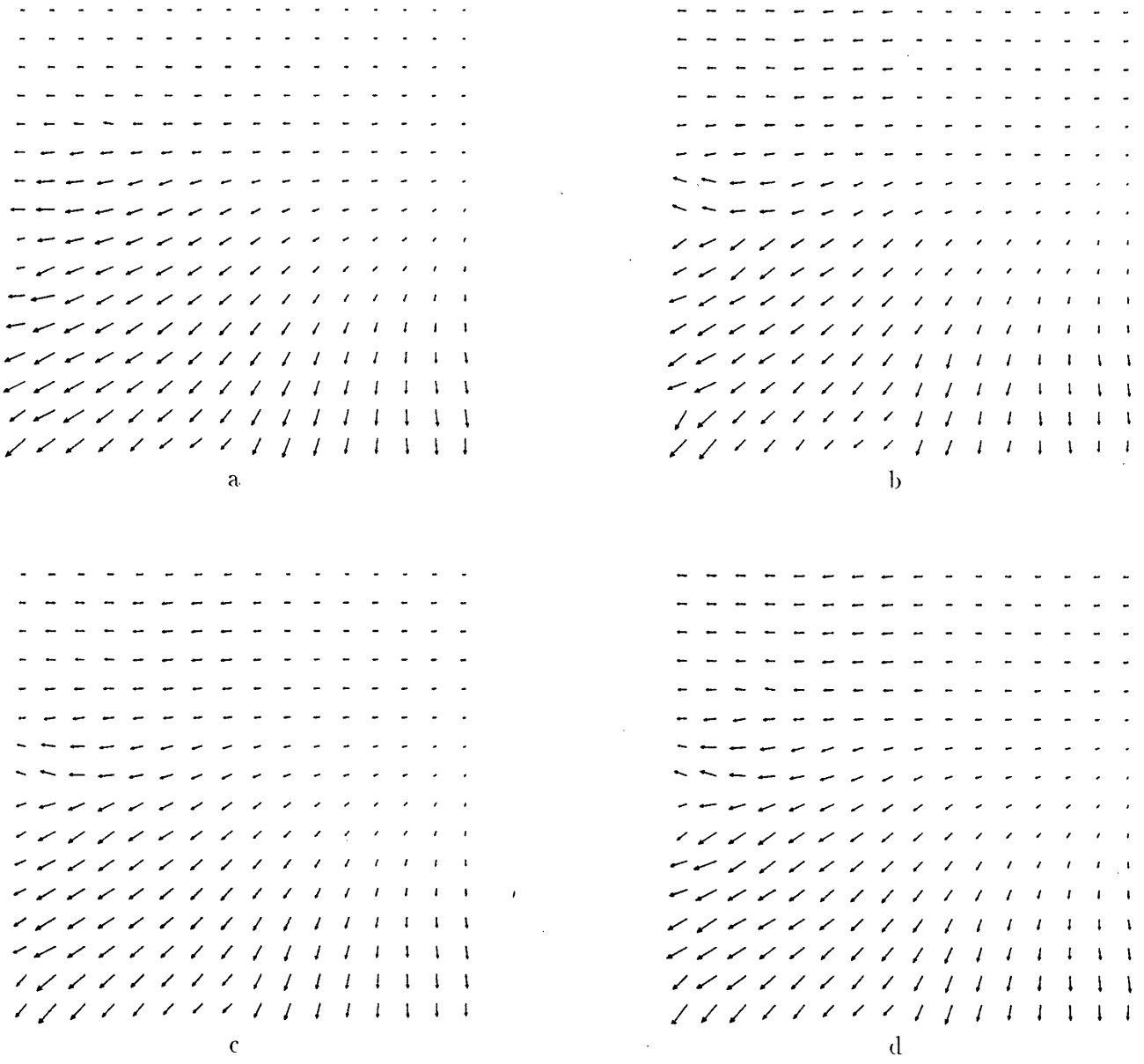


Figure 13: Yosemite sequence flow estimates. (a) Smoothness constraint estimates computed using 100 iterations of SOR. (b) Multiscale Regularization (MR) estimates. (c) Post-filtered MR estimates and (d) Estimates produced by using MR estimates as initial condition for SOR algorithm.

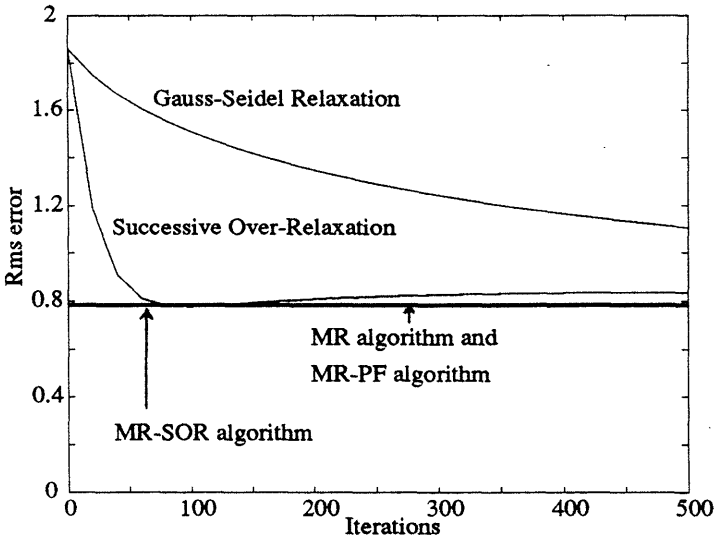


Figure 14: Rms Error Comparison of MR, MR-PF, MR-SOR, SOR and Gauss-Seidel (GS) algorithm flow estimates for the yosemite sequence.

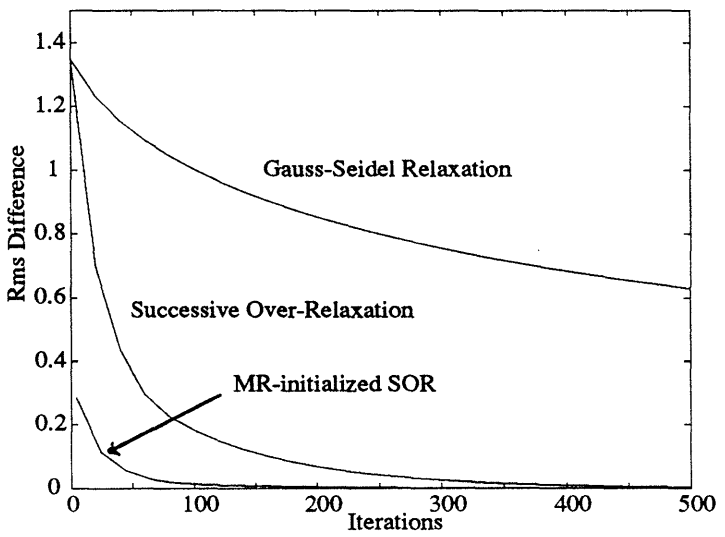


Figure 15: Rms Difference Comparison illustrates how the MR-initialized SOR, SOR and GS algorithms converge to the smoothness constraint solution (Yosemite sequence).

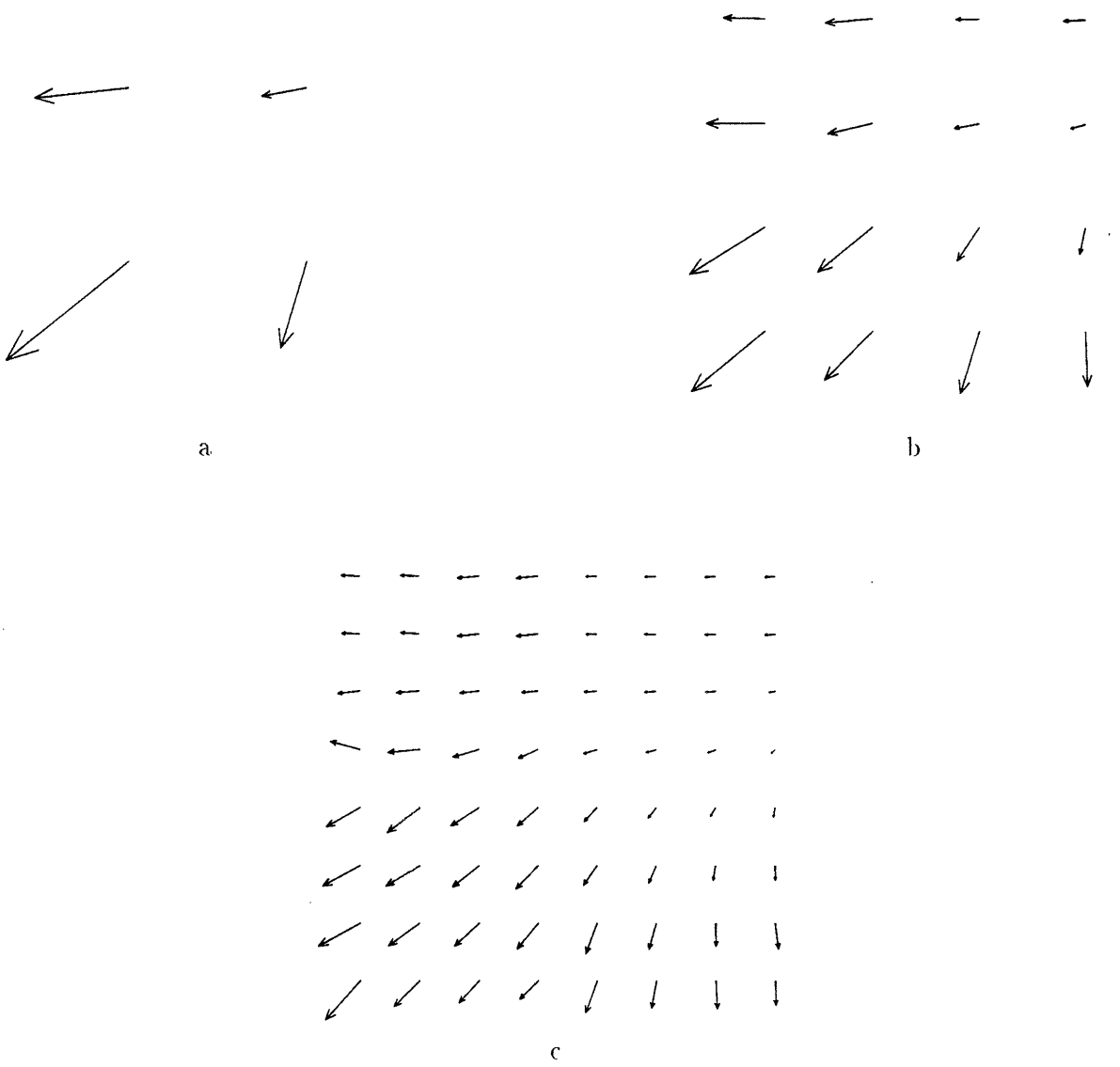


Figure 16: Multiscale Regularization flow estimates at the (a) first. (b) second and (c) third scales.

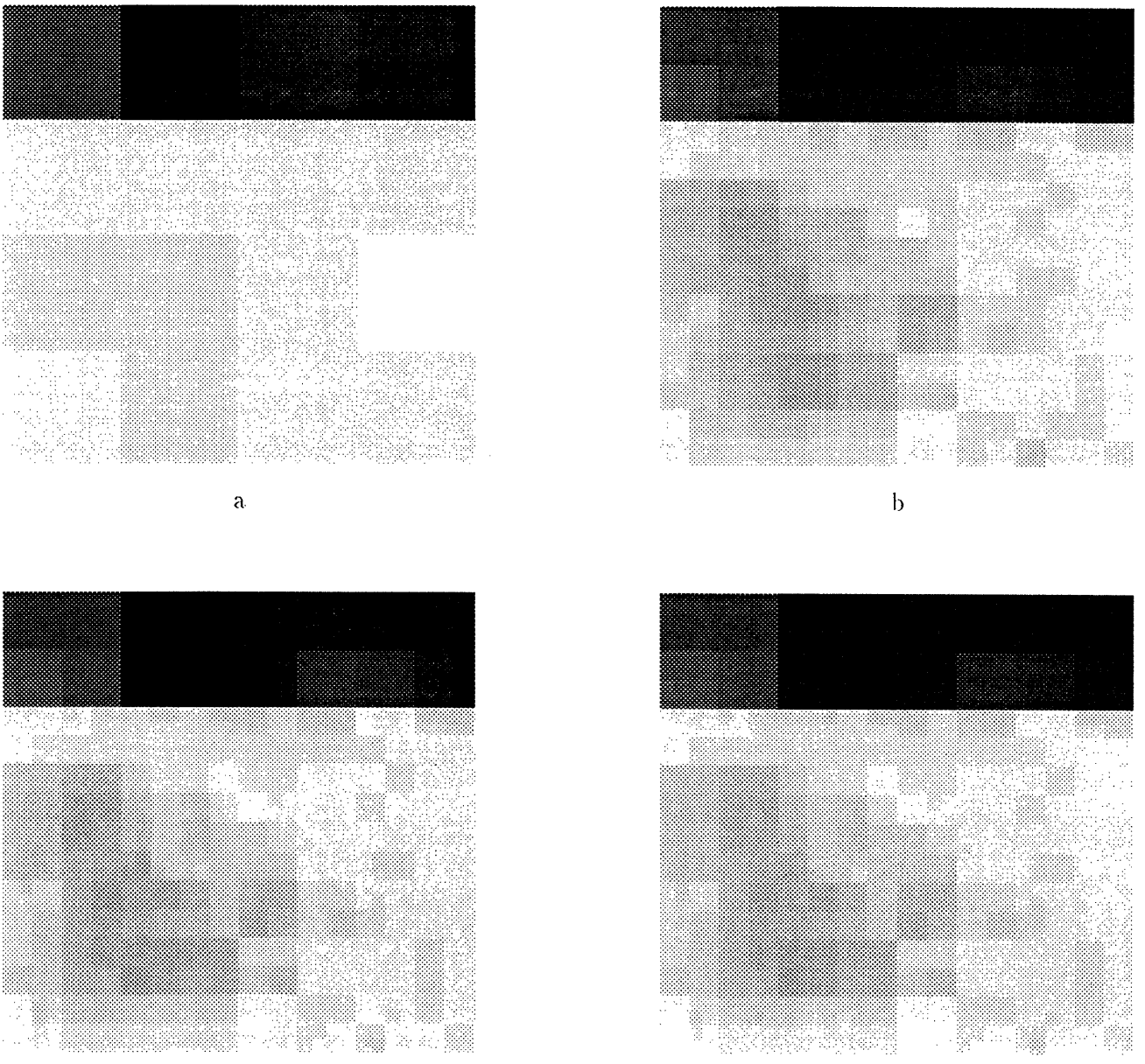


Figure 17: Multiscale Regularization error covariance at the (a) second, (b) fourth, (c) sixth and (d) finest scales.

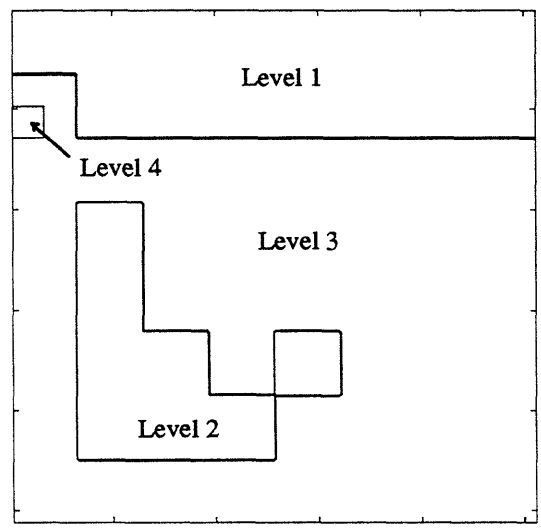


Figure 18: Map depicting the optimal resolution for representing the optical flow field as a function of lattice site. Note that the optical flow field is represented at a coarser level in the quadtree in regions where there is no gradient information (at the top). It is also represented at a coarser level along the face of the mountain, where there is little gradient information parallel to the striations.

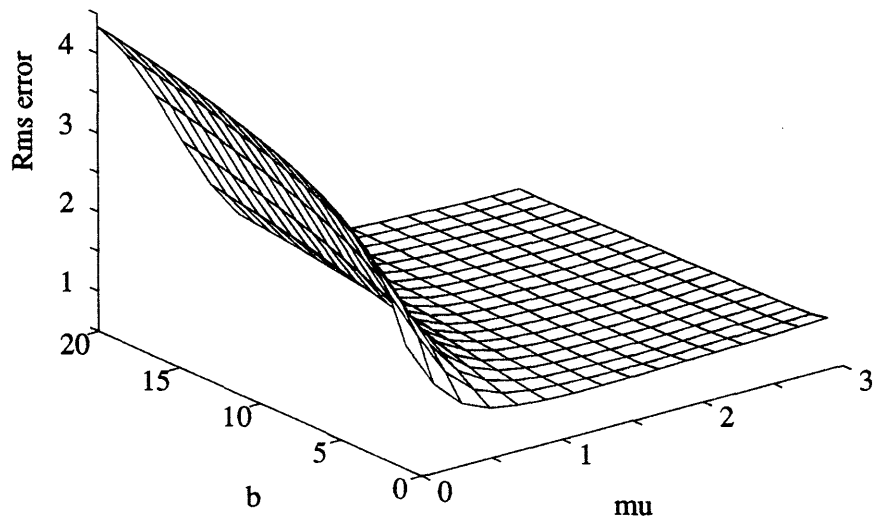


Figure 19: Multiscale Regularization rms error sensitivity to the parameters b and μ (Yosemite sequence).

Figure 20: First frame of Moving vehicle sequence.

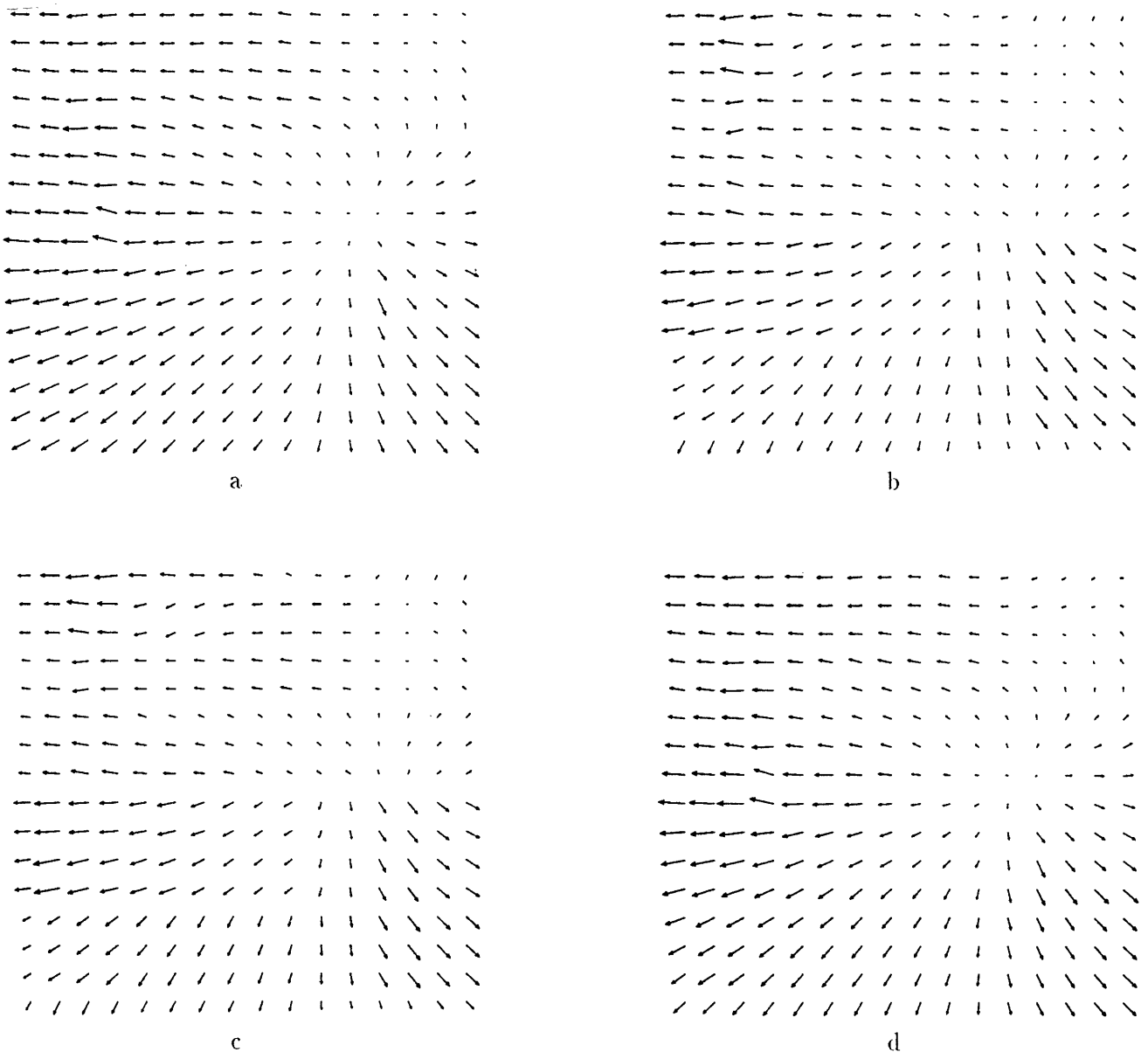


Figure 21: Moving vehicle sequence flow estimates. (a) Smoothness constraint estimates computed using 300 iterations of SOR, (b) Multiscale Regularization (MR) estimates. (c) Post-filtered MR estimates and (d) Estimates produced by using MR estimates as initial condition for SOR algorithm.

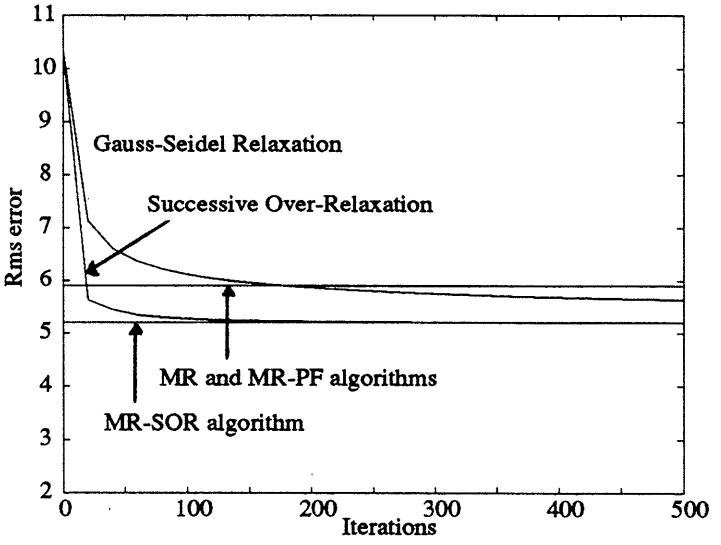


Figure 22: Rms Error Comparison of MR, SOR and Gauss-Seidel (GS) algorithm flow estimates for the Moving vehicle sequence.

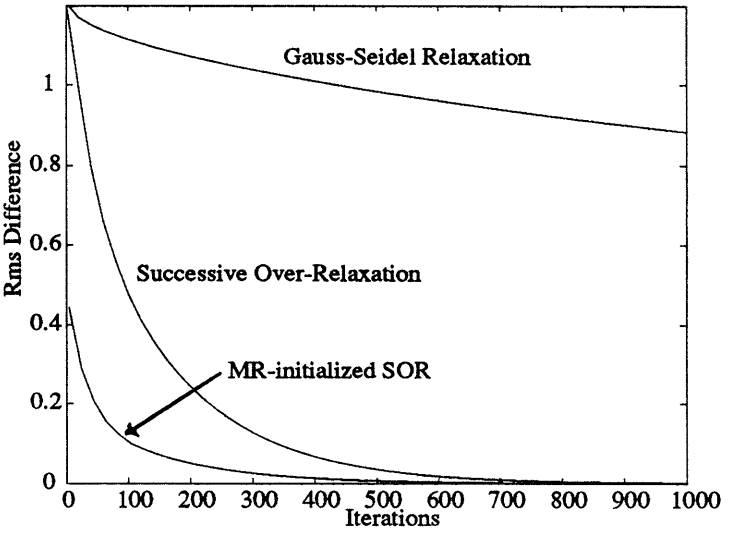
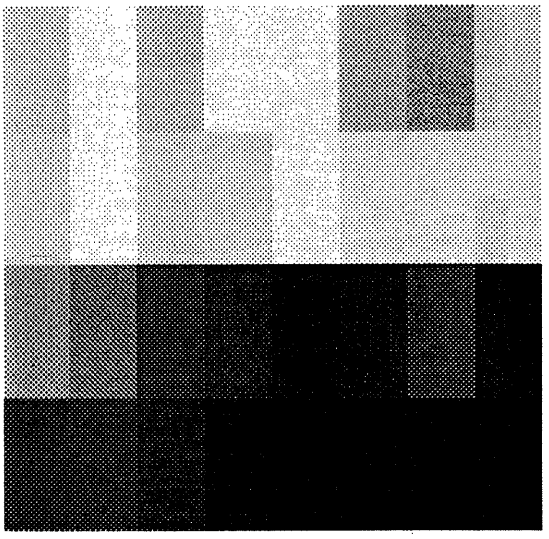
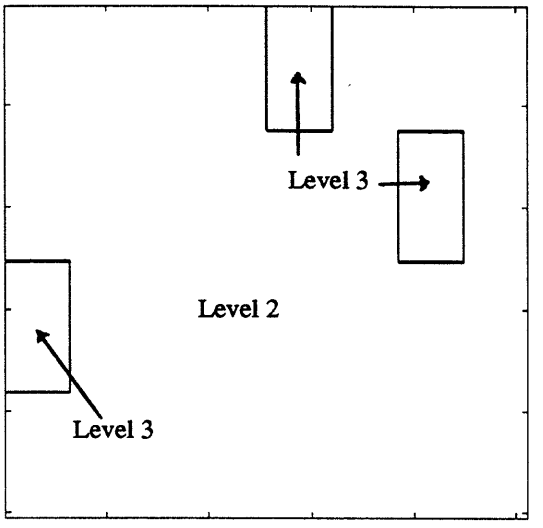


Figure 23: Rms Difference Comparison illustrates how the MR initialized SOR, SOR and GS algorithms converge to the smoothness constraint solution (Moving vehicle sequence).



a



b

Figure 24: (a) Multiscale Regularization error covariance at the finest scale and (b) Map illustrating the optimal representation resolution for the Moving vehicle sequence optical flow estimates.

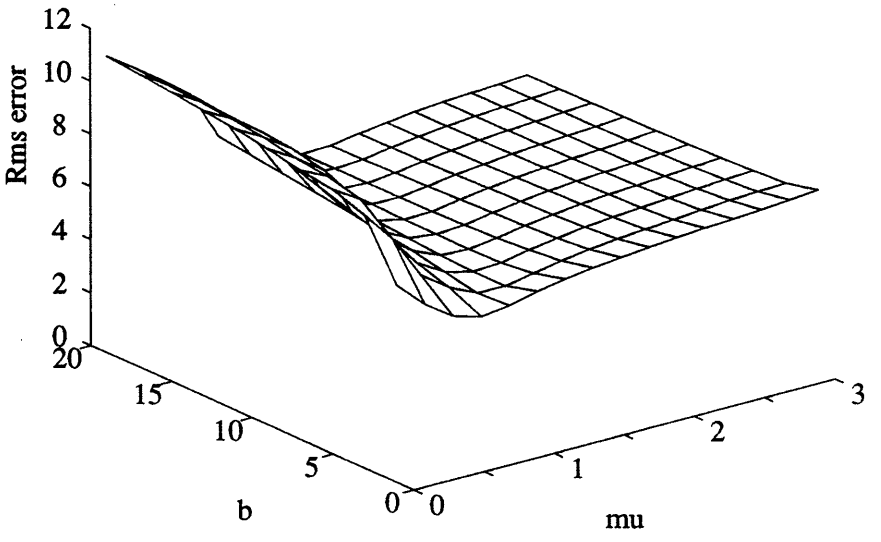


Figure 25: Multiscale Regularization rms error sensitivity to the parameters b and μ (Moving vehicle sequence).

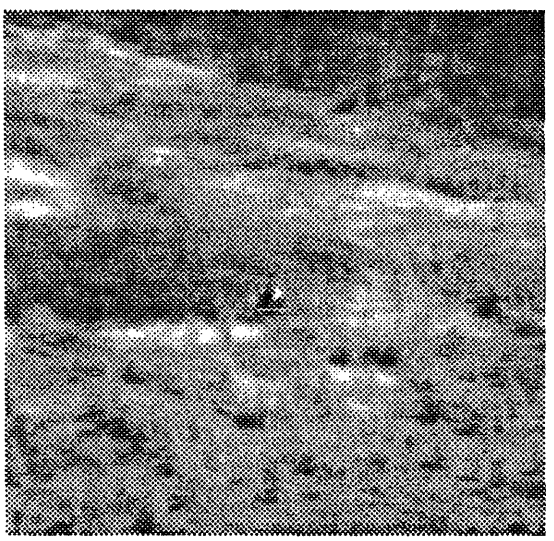


Figure 26: First frame of Chopper sequence.

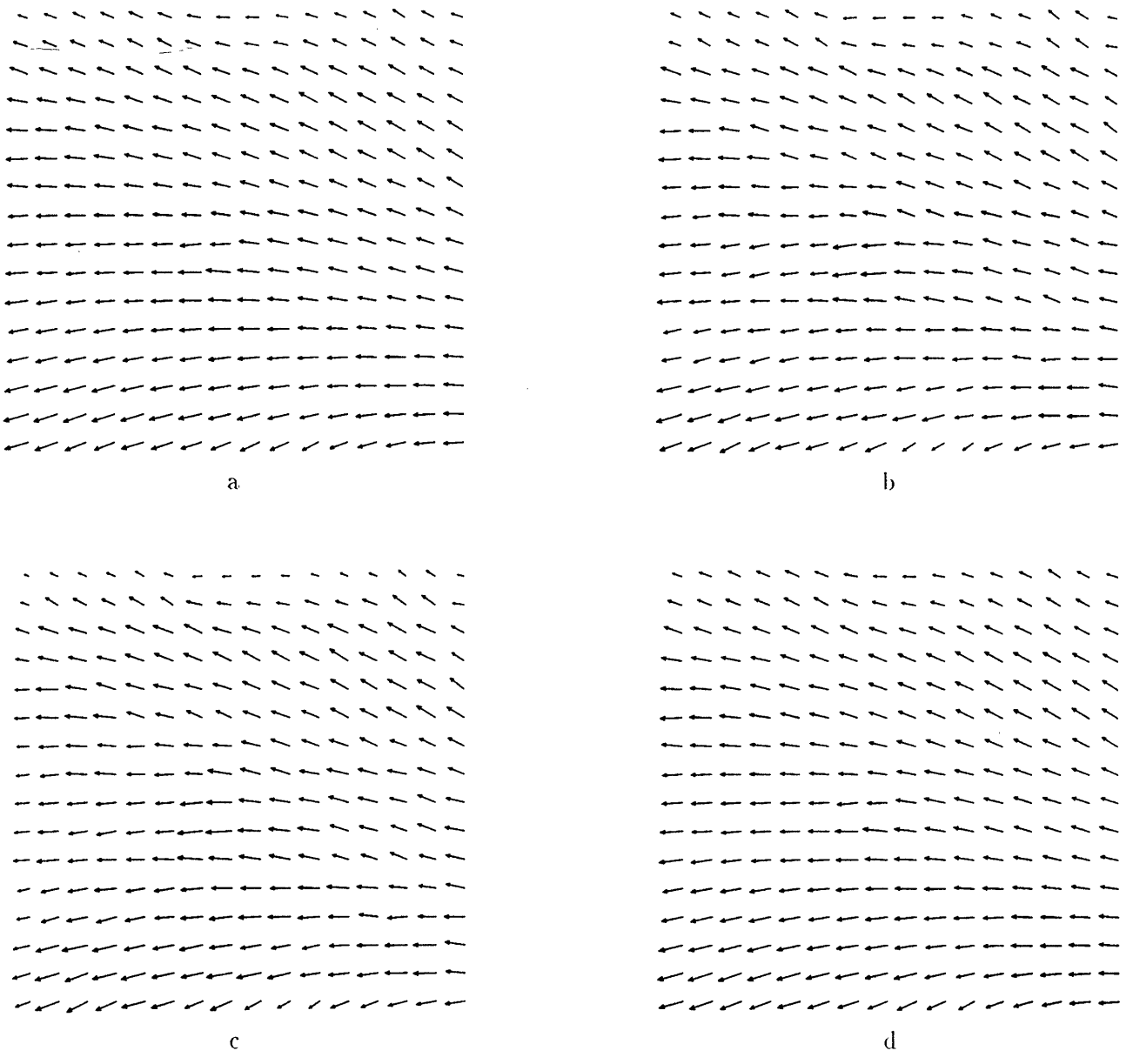


Figure 27: Chopper sequence flow estimates. (a) Smoothness constraint estimates computed using 200 iterations of SOR, (b) Multiscale Regularization (MR) estimates, (c) Post-filtered MR estimates and (d) Estimates produced by using MR estimates as initial condition for SOR algorithm.

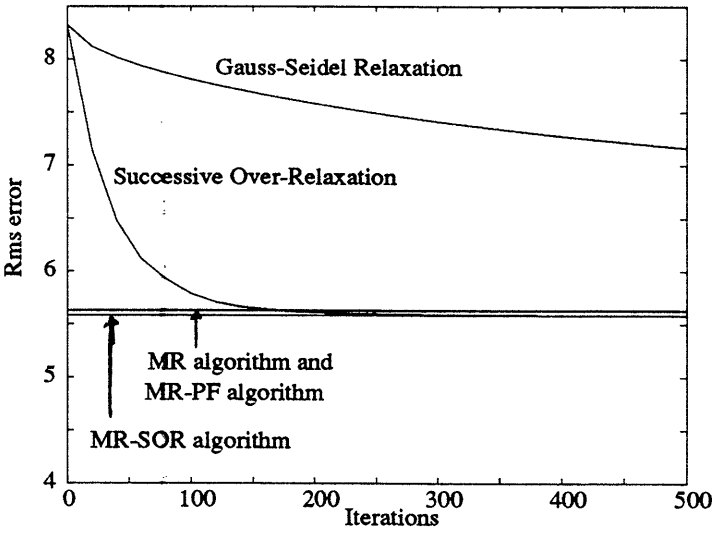


Figure 28: Rms Error Comparison of MR, SOR and Gauss-Seidel (GS) algorithm flow estimates for the Chopper sequence.

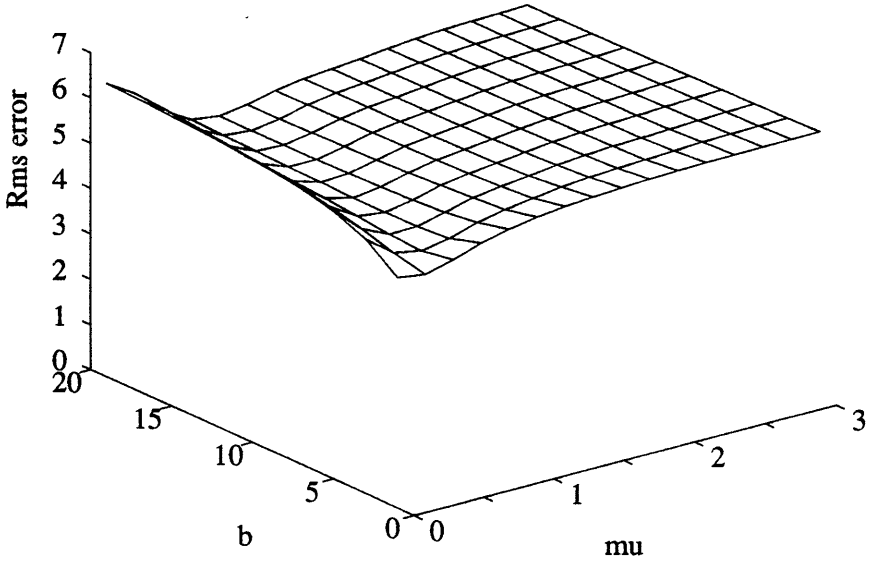


Figure 29: Multiscale Regularization rms error sensitivity to the parameters b and μ (Chopper sequence).

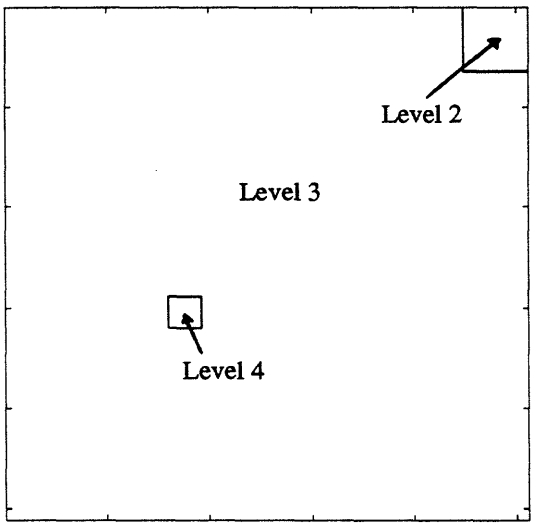


Figure 30: Map illustrating the optimal resolution for the Chopper sequence optical flow estimates.

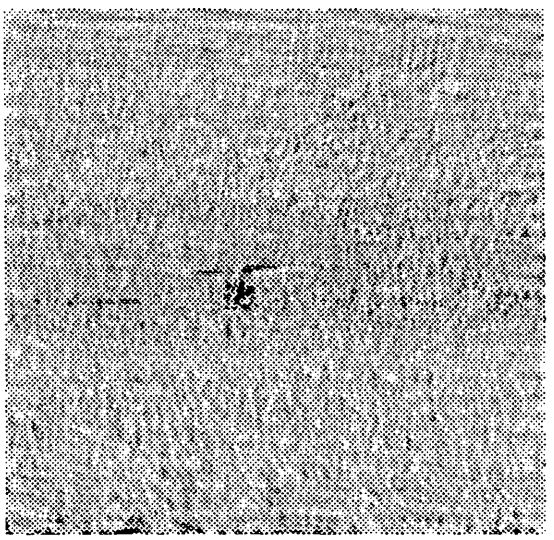


Figure 31: The smoothing filter residuals shown above can be used to develop adaptive algorithms for the motion-based object detection.

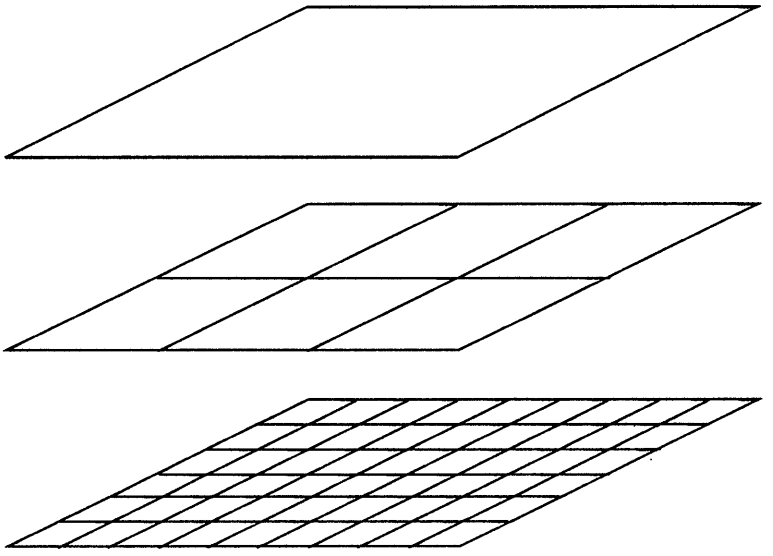


Figure 32: Non-homogeneous tree structure for lattices which are not square. The grid structure is a simple extension of the quadtree structure in that it allows for varying numbers of "offspring" from each parent. The figure illustrates a hierarchy of grids for a 6×9 lattice.